From Active-Site Mapping to Lead Discovery using Fragment-based Approaches on the Aspartic protease Endothiapepsin

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Pharmazie

der PHILIPPS-UNIVERSITÄT MARBURG

vorgelegt von

Nedyalka Radeva

aus Kardzhali

Marburg/Lahn 2016

Von Fachbereich Pharmazie der Philipps-Universität Marburg

als Dissertation eingereicht am: 01.12.2016

Erstgutachter: Prof. Dr. Gerhard Klebe

Zweitgutachter: Prof. Dr. Andreas Heine

Tag der mündlichen Prüfung: 25.01.2017

Hochschulkennziffer: 1180

Die Untersuchungen zum vorliegenden Thema wurden auf Anregung von Prof. Dr. Gerhard Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von August 2012 bis November 2016 durchgeführt.

Nicht die Neugier entscheidet über unser Wissen, sondern die Bereitschaft Altes aufzugeben.

Meiner Familie

This thesis resulted in several scientific contributions:

I. Journal Articles:

- Johannes Schiebel, **Nedyalka Radeva**, Helene Köster, Alexander Metz, Timo Krotzky, Maren Kuhnert, Wibke E. Diederich, Andreas Heine, Lars Neumann, Cedric Atmanene, Dominique Roecklin, Valérie Vivat-Hannah, Jean-Paul Renaud, Robert Meinecke, Nina Schlinck, Astrid Sitte, Franziska Popp, Markus Zeeb, Gerhard Klebe. One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* **2015**, *10*, 1511–1521.

- **Nedyalka Radeva**, Johannes Schiebel, Xiaojie Wang, Stefan G. Krimmer, Kann Fu, Martin Stieler, Frederik R. Ehrmann, Alexander Metz, Thomas Rickmeyer, Michael Betz, Johan Winquist, Ah Young Park, Franziska U. Huschmann, Manfred S. Weiss, Uwe Mueller, Andreas Heine, Gerhard Klebe. Active-Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads to Address a Catalytic Dyad. *J.Med.Chem.* **2016**, *59*, 9743–9759.

- **Nedyalka Radeva**, Stefan G. Krimmer, Martin Stieler, Kan Fu, Xiaojie Wang, Frederik R. Ehrmann, Alexander Metz, Franziska U. Huschmann, Manfred S. Weiss, Uwe Müller, Johannes Schiebel, Andreas Heine, Gerhard Klebe. Experimental Active-Site Mapping by Fragments – Hot Spots Remote from the Catalytic Center of Endothiapepsin. *J.Med.Chem.* **2016**, *59*, 7561–7575.

- Johannes Schiebel, **Nedyalka Radeva**, Stefan G. Krimmer, Xiaojie Wang, Martin Stieler, Frederik R. Ehrmann, Kan Fu, Alexander Metz, Franziska U. Huschmann, Manfred S. Weiss, Uwe Müller, Andreas Heine, Gerhard Klebe. Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chem.Biol.* **2016**, *11*, 1693–1701.

- Johannes Schiebel, Stefan G. Krimmer, Karine Röwer, Anna Knörlein, Xiaojie Wang, Ah Young Park, Martin Stieler, Frederik R. Ehrmann, Kan Fu, **Nedyalka Radeva**, Michael Krug, Franziska U. Huschmann, Steffen Glöckner, Manfred S. Weiss, Uwe Mueller, Gerhard Klebe, Andreas Heine. High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* **2016**, *24*, 1398–1409.

- Milon Mondal, **Nedyalka Radeva**, Helene Köster, Ah Young Park, Constantinos Potamitis, Maria Zervou, Gerhard Klebe, Anna K. H. Hirsch. Structure-based design of inhibitors of the aspartic protease endothiapepsin by exploiting dynamic combinatorial chemistry. *Angew.Chem.Int.Ed.* **2014**, *53*, 3259–3263.

- Alwin M. Hartman, Milon Mondal. **Nedyalka Radeva**, Gerhard Klebe, Anna K. H. Hirsch. Structure-based optimization of inhibitors of the aspartic protease endothiapepsin. *Int.J.Mol.Sci.* **2015**, *16*, 19184-19194.

- Milon Mondal, **Nedyalka Radeva**, Hugo Fanlo-Virgós, Sijbren Otto, Gerhard Klebe, Anna K. H. Hirsch. Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin: Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin: Fragment-Based Drug Design Facilitated by Dynamic Combinatorial Chemistry. *Angew.Chem.Int.Ed.* **2016**, *55*, 9422–9426.

- Jonathan Cramer, Johannes Schiebel, Tobias Wulsdorf, Kristof Grohe, Eszter Eva Najbauer, Frederik R. Ehrmann, **Nedyalka Radeva**, Nina Zitzer, Dr. Uwe Linne, Prof. Dr. Rasmus Linser, Prof. Dr. Andreas Heine, Prof. Dr. Gerhard Klebe. Watch out for the Red Herring: Surprising Reactivity of a Fragment Results in Biological Activity. *Angew.Chem.Int.Ed.*, DOI: 10.1002/anie.201609824R1 and 10.1002/ange.201609824R1

II. Oral presentations:

- Design of Novel Aspartic Protease Inhibitors Exploiting Dynamic Combinatorial Chemistry, *22nd Annual Conference of the German Crystallographic Society (DGK)*, Berlin **2014**
- Design of Novel Aspartic Protease Inhibitors Exploiting Dynamic Combinatorial Chemistry, *International School of Crystallography, 47th Course*, Erice **2014**

Table of content

1. Introduction

1. Introduction

1.1 Structure-Based Drug Design

Structure-based drug design (SBDD) is an important approach for accelerating the drug-discovery process.[1,2] The basic principles rely on designing and optimizing a small molecule in a way that it fills optimally the active site of a target protein while disturbing the function of the latter and thereby achieving a therapeutic effect. The understanding of the molecular interactions in protein-ligand complexes is thus key. Therefore, the three-dimensional structure of a target protein is needed, which is usually determined by X-ray crystallography or NMR spectroscopy. If no structure of the target exists, a homology model can be calculated derived from closely related proteins of the same family.

SBDD is an iterative process, where after an appropriate target has been selected, a large number of library compounds are screened for their biological activity against this target through different high-throughput screening (HTS) techniques. Computational approaches are used as a complement to HTS techniques.[3] Once a ligand with suitable affinity is found, the structure of the corresponding protein-ligand complex is determined. This information usually serves as a starting point for further optimization procedures, such as the synthesis of more appropriate ligands in order to optimize the affinity contribution followed by testing the biological activity of the new compounds.

The affinity or the interaction of a ligand with its target known from the three-dimensional structure is the prerequisite for considering the ligand for further optimization. The number of protein structures deposited in the protein data bank (PDB) has steadily been increasing during the last twenty years (Fig. 1.1, status from September 2016).[4] Nevertheless, there are still many other criteria that have to be fulfilled in parallel in order to bring a promising ligand into preclinical studies. In latter stages of optimization for instance, the so-called ADME parameters (**A**dsorption, **D**istribution, **M**etabolism, and **E**xcretion) are addressed.

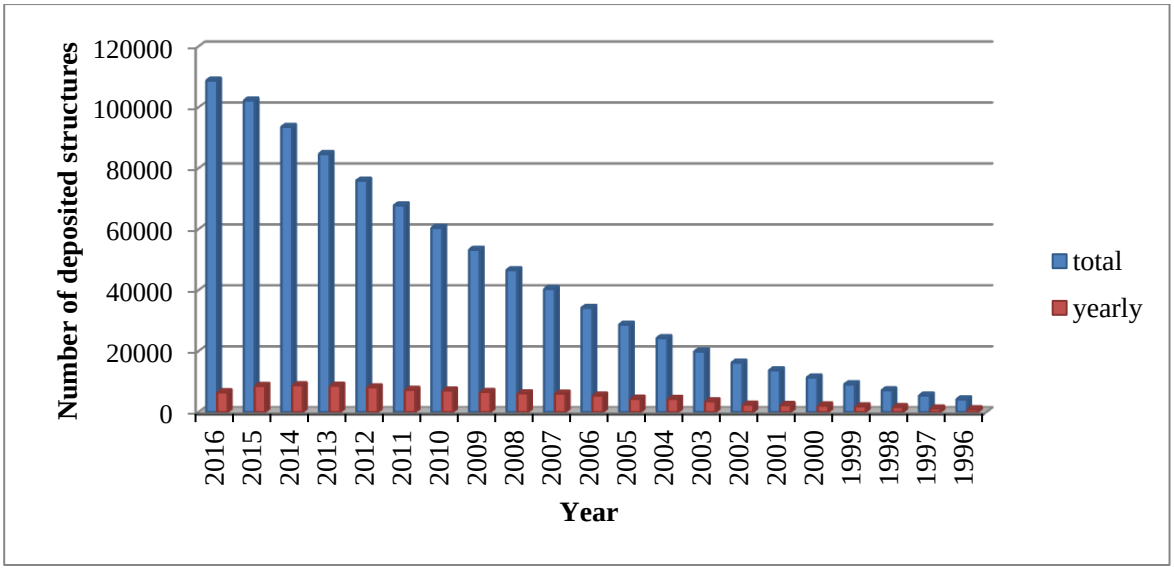


***Fig. 1.1**. Structures determined by X-ray crystallography and deposited in the PDB during the last twenty years.*

1.2 Fragment-Based Drug Discovery

The pharmaceutical industry is currently dealing with unprecedented productivity challenges due to the steadily increasing regulatory requirements and the dropping monetary turnover owing to patent expirations.[5] The associated costs with a successful launch of a new lead to the market come to approximately $1.8 billion and it takes on an average 1.7 years to pass through each stage of development as depicted in Fig. 1.2.[5,6] Some twenty years ago HTS was the leading drug-discovery method, in particular in the pharmaceutical industry.[7] The main advantage of this approach is the possibility of screening thousands of molecules within a short time frame. Nevertheless, the expansion of the chemical space with all possible drug-like molecules estimated at ~10^{63} represents a huge challenge making even a library with millions of compounds look rather poor.[8] Furthermore, although HTS hits often exhibit appropriate affinity contributions in the low µM to the high nM ranges, they have disappointing low ligand efficiency (LE; eq. 1) as the number of atoms involved in direct interactions with key pocket residues is rather minor (Fig. 1.3).[9,10]

$$\mathrm{LE} = \frac{-2.303\,RT}{HA}.\,log\,K_d \qquad [1]$$

Fragment-based drug discovery (FBDD) on the other hand, meanwhile widely used in academia and industry, has emerged more and more as a novel paradigm in lead discovery over the past two decades, as it represents a great alternative to common HTS techniques.[11–13] The principles of FBDD originate from the early eighties based on the pioneering work of Jencks describing that the Gibbs free energy of binding (ΔG^0) of a larger molecule A-B is the result of the combination of the free energies of binding of two or more fragment molecules.[14] But how are fragments actually defined? For this purpose, Congreve *et al.* at Astex introduced the so-called rule of three (RO3)[15] alluding to criteria similar to the Lipinski's rule of five (RO5)[16]. These describe the properties of a fragment in the following way: molecular weight (MW) < 300 Da, clogP (o/w) ≤ 3, number of H-bond donors ≤ 3, number of H-bond acceptors ≤ 3, and number of rotational bonds ≤ 3. However, as many screening methods fail to identify fragment binding owing to their weak affinity and not least due to solubility issues, researchers tend to use even smaller fragments restricting the number of heavy atoms with an upper limit of 17, whereby the lipophilicity is likely to be lower while improving the fragment solubility in water.[17,18]

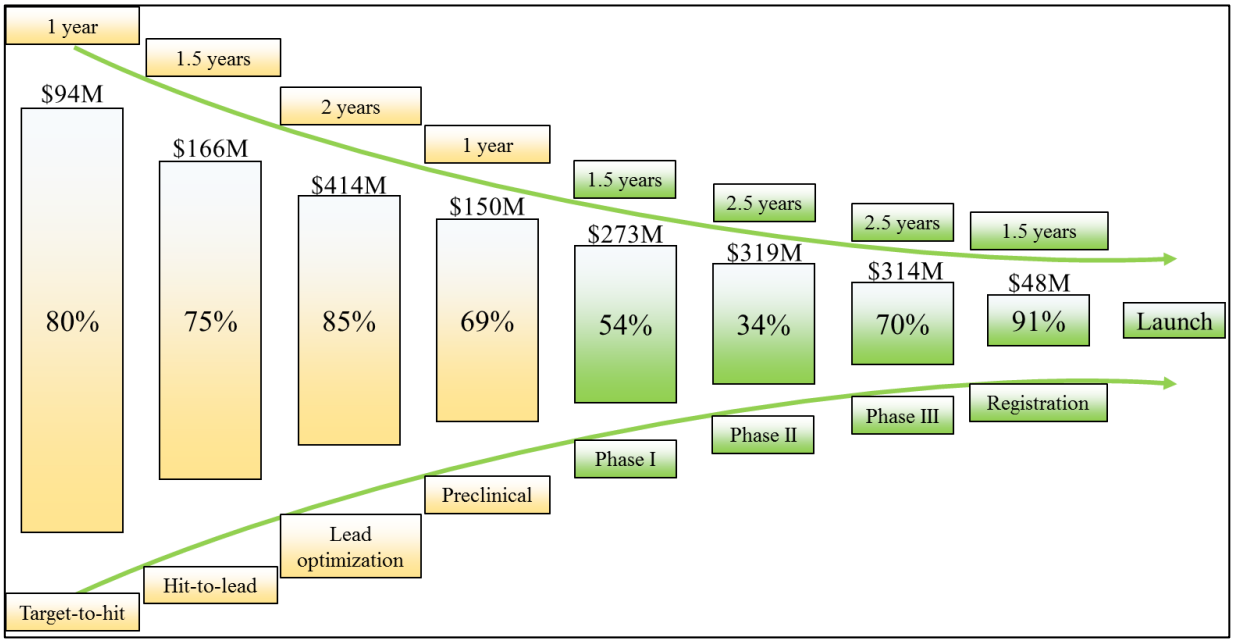


***Fig. 1.2**. Duration time and costs for each individual stage of development, required for launching a new lead to the market. The percentages depicted in the bars indicate the probability of successful passage from one state of development to another. The figure is adapted from S. M. Paul et. al. Nat. Rev. Drug Discovery 2010, 9, 203–214.*

This technique offers many attractive features compared to HTS techniques. For instance, a much smaller number of compounds is usually screened owing to the higher propensity of fragments to match the binding site of a target.[19]

Furthermore, although fragments have low MW ~ 150−300 Da they exhibit high LE as a higher number of atoms is involved in direct interactions with key pocket residues. Moreover, the low MW of the fragments is a huge advantage when it comes to improving the binding affinity and selectivity as chemical functionalities can be added to the basic molecular scaffold in iterative manner maintaining the range for lead-likeness (Fig. 1.4).[20] The technique holds great promise as there are currently two drugs emerging from FBDD approaches on the market and many more are currently in clinical trials (Table 1). The first one (Vemurafenib, *Zelboraf* ®) was approved by the US Food and Drug Administration (FDA) as early as 2011 for the treatment of late-stage melanoma. The development of *Zelboraf* ® started as a screening campaign of 20 000 compounds for their ability to inhibit a number of kinases.[21] The additionally performed crystallographic analysis with ~1% of the initial compounds revealed the structure of an azaindole bound to the Pim1 kinase, the optimization of which led to the structure of Vemurafenib. The second one, Venetoclax (*Venclexta*®), was approved by the FDA in April 2016 for the treatment of chronic lymphotic leukemia.[22]

However, as already mentioned above, the methods usually applied to identify fragment binding are limited in their capability to achieve high hit rates, often owed to the low solubility of fragments in aqueous solutions.[23] In the following, the basic principles of the applied techniques and the properties of the studied systems in this thesis will be presented.

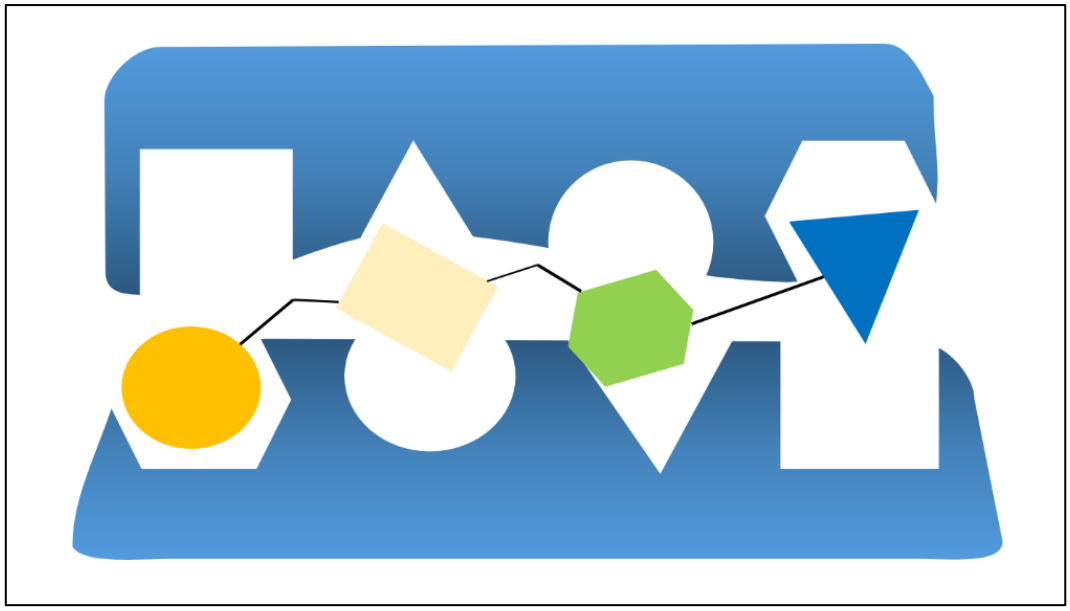


***Fig. 1.3**. Example of bad quality interactions of a HTS hit with its target protein. The figure is adapted from D. C. Rees et. al. Nat. Rev. Drug Discovery 2004, 3, 660–672.*

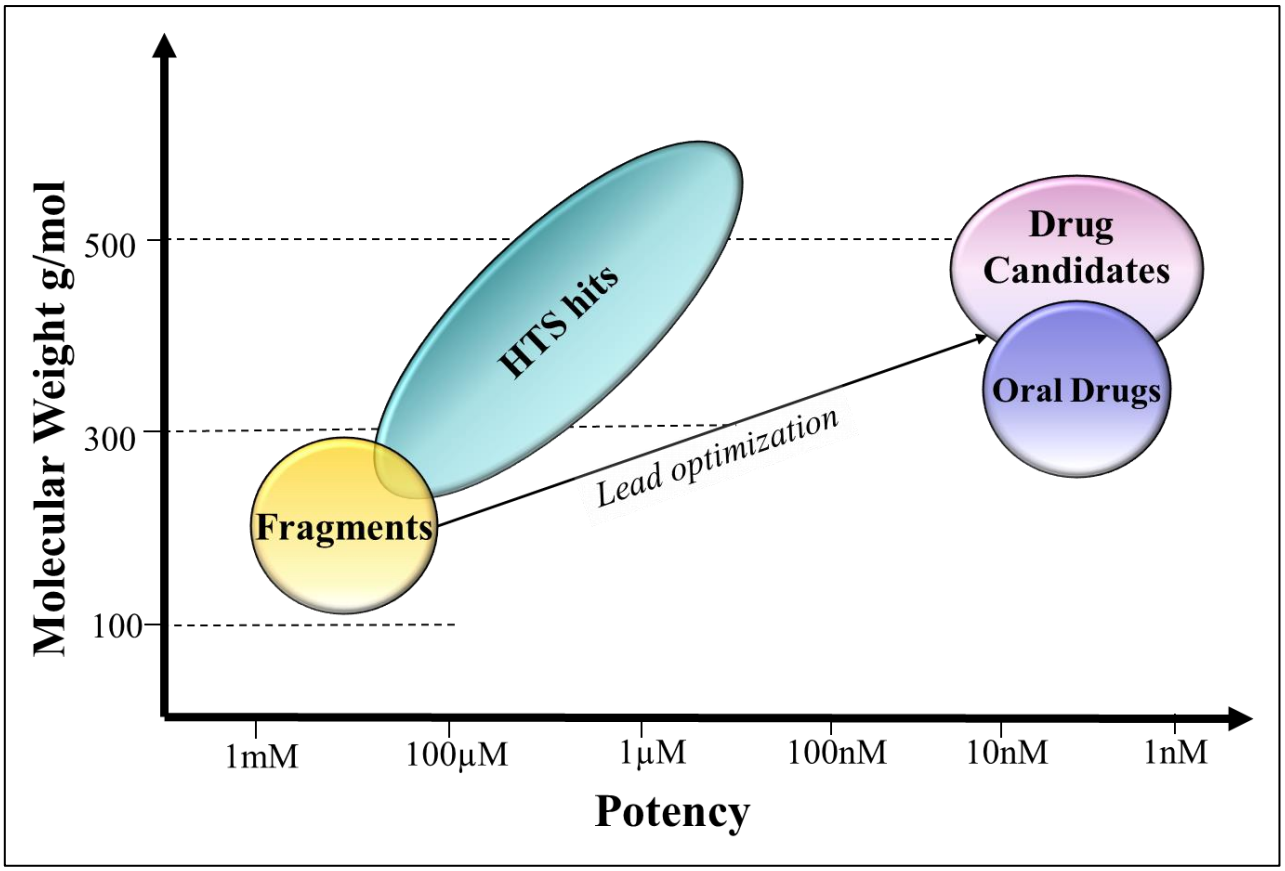


***Fig. 1.4**. Schematic illustration of the relation between MW and affinity of fragment and HTS-hits. The figure is adapted from D. C. Rees et. al. Nat. Rev. Drug Discovery 2004, 3, 660–672.*

Table 1.1. *Examples of clinical candidates emerged from fragments with entries on the US National Institutes of Health website ClinicalTrials.gov.*[24]

Compound Name	**Developer**	**Therapeutic area**	**Highest phase reached**	**Ref.**[25] **or Clinical trial identifier**
B cell lymphoma (BCL-2) family				
Venetoclax (ABT-199)	AbbVie	Cancer	Pre-registration	NCT01682616
Mouse double minute 2 homologe (MDM2)-p53				
Idasanutlin (RO5503781)	Roche	Cancer	Phase III	NCT02545283
AMG232	Amgen	Cancer	Phase I/II	NCT02110355
CGM097	Novartis	Cancer	Phase I	NCT01760525
DS-3032b	Daiichi Sankyo	Cancer	Phase I	NCT01877382
ALRN-6924	Aileron	Cancer	Phase I	NCT02264613
MK-8242	Merck	Cancer	Phase I	NCT01463696
JNJ-26854165	Johnson&Johnson	Cancer	Phase I	NCT00676910
Leukocyte function-associated molecule 1 (LFA1)-Intracellular adhesion molecule 1 (ICAM1)				
Lifitegrast (SAR1118)	Shire (SARcode bioscience)	Dry eye	Pre-registration	NCT01743729
aIIbß3				
Tirofiban	Merck	Cardiovascular	Approved	NCT01109134
a4ß1				
Carotegrast methyl (AJM300)	Ajinomoto	Ulcerative colitis	Phase III	Cox *et al* 2010
Inhibitor of apoptosis protein (IAP)				
AT-406 (Debio-1143)	Ascenta Therapeutics/Debiopharm	Cancer	Phase II	NCT02022098
LCL-161	Novartis	Cancer	Phase II	NCT01955434
Birinapant (TL32711)	Tetralogic Pharma	Cancer	Phase I/II	NCT02587962
ASTX-660	Astex therapeutics	Cancer	Phase I/II	NCT02503423
AEG40826 (HGS1029)	Aegera and Human Genome Sciences	Cancer	Phase I	NCT00708006
CUDC-427 (GDC-0917)	Curis (Genentech)	Cancer	Phase I	NCT01908413
Bromodomain family				
Apabezalone (RVX-000222, RVX-208)	Resverlogix	Cardiovascular	Phase III	NCT02586155
GSK525762	GlaxoSmithKline	Cancer	Phase I	NCT01587703
CPI-0610	Constellation Pharmaceuticals	Cancer	Phase I	NCT02157636
TEN-010	Tensha Therapeutics	Cancer	Phase I	NCT02308761
OTX015	Oncoethix (Merck)	Cancer	Phase I	NCT02259114

1.3 Experimental Methods used for Identifying Fragment Binding

1.3.1 X-ray Crystallography

X-ray crystallography is a reliable hit-identification method revealing more and more its potential as fragment screening technique.[26,27] It is based on the analysis of the reflection pattern of a crystal, in which the atomic arrangement causes a beam of incident electromagnetic radiation (X-ray) to diffract. For the optimal usage of X-ray crystallography, a robust crystallization system with well diffracting crystals is required. For many targets, such as membrane-bound proteins, the usage of this approach for screening fragment libraries is presently still rather unsuitable due to difficulties of growing sufficient crystals. Nevertheless, X-ray crystallography is key in the process of hit-to-lead optimization, thus making its application as primary screening method even more attractive as a hit comes directly with a crystal structure. A huge advantage compared to other screening techniques is the high compound concentration, which can be applied during soaking experiments (~ 100mM) as compound precipitation and thus deposit is not an issue unlike for other screening techniques. Even though the high concentration often leads to the appearance of bound ligands with otherwise very low affinity, which would fail as hits in any other screening technique, the structural information can be very useful with regard to optimal extension of a basic molecular scaffold within the protein binding site. Moreover, the method is very effective in avoiding false positives since an observed electron density in the crystal structure clearly indicates a bound compound.

1.3.1.1 Bragg's Law

The lattice planes of regularly arranged atoms in a crystal scatter an regular array of spherical waves. The reflected beam waves often cancel one another out through destructive interference (Fig. 1.5, A) but in a few specific directions they add up only if the Bragg's law is fulfilled (Fig. 1.5, B). This is the case when the integer *n* multiple of the wavelength λ of the incoming beam is equal to the path distance $2d\sin\Theta$ of the reflected beam (eq. 2).

$$2d \cdot \sin\Theta = n \cdot \lambda \qquad [2]$$

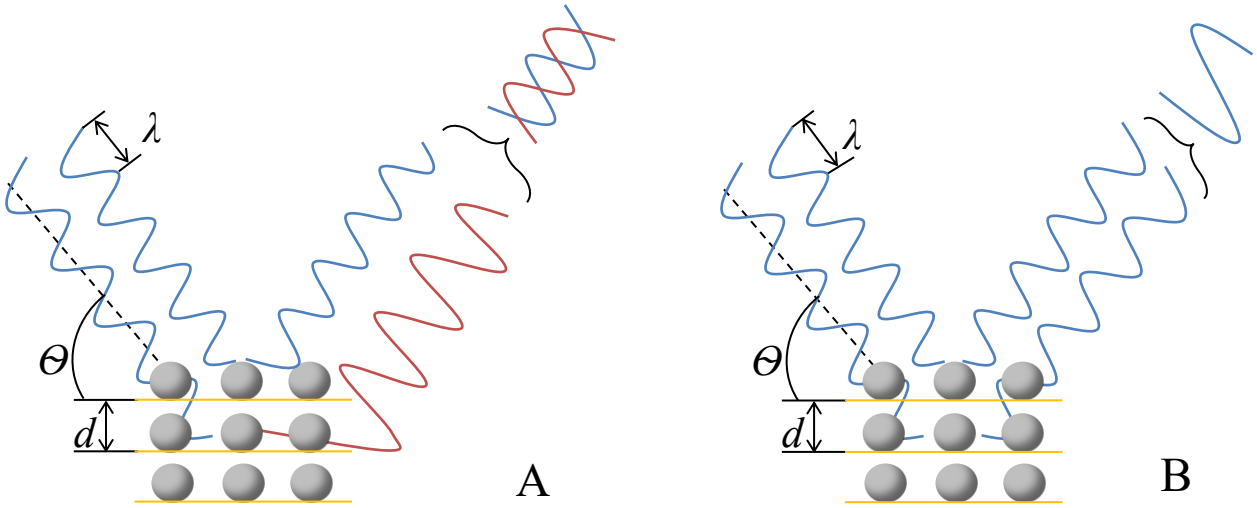


***Fig. 1.5**. Bragg diffraction, according to which the phase shift causes destructive (A) or constructive (B) interferences.*

The crystal lattice reflects the incoming beam, in a very similar fashion as a mirror, and we obtain the diffraction pattern created by the electrons of the atoms (Fig. 1.6, A). The diffraction pattern is reciprocal to the crystal lattice. It occurs only when series of parallel crystal planes intersect the so-called Ewald sphere, resulting in reflection spots on the detector (Fig. 1.6, B). Every single reflection spot includes the sum of the scattering of all atoms within the unit cell.

During the experiment the information concerning the phases of every diffracted wave is lost. The diffraction data provide only the amplitude of the so-called Fourier transform of the electron density in the unit cell.

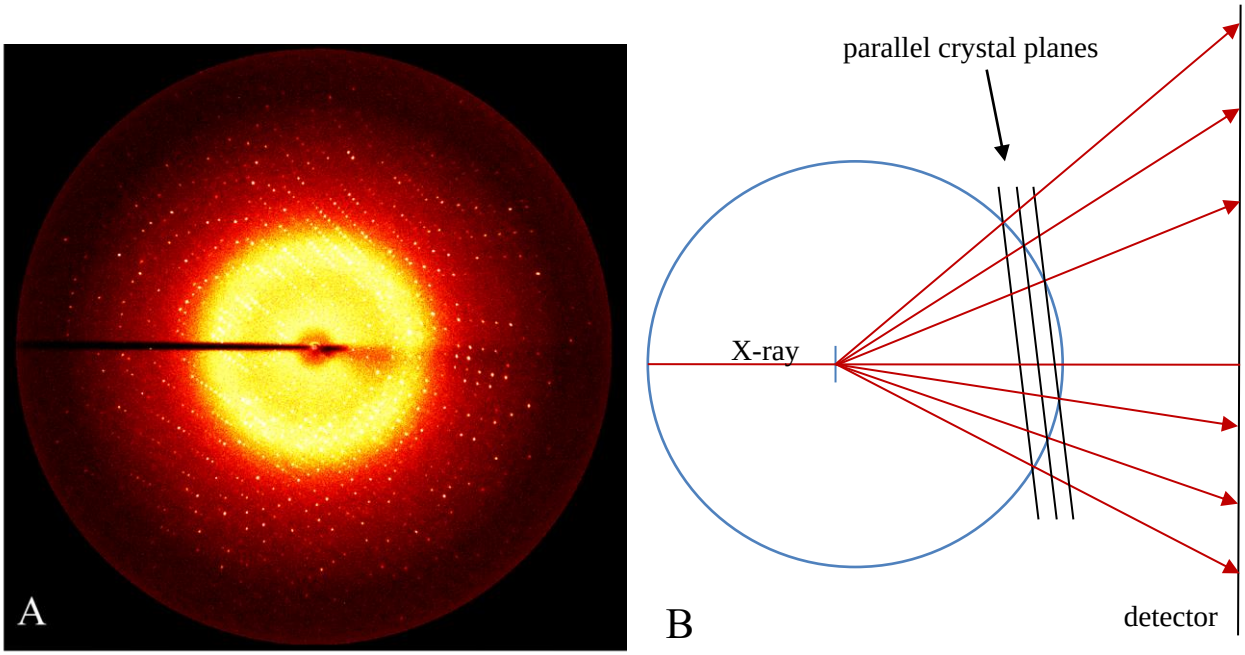


***Fig. 1.6**.* A*: Diffraction pattern of an EP crystal.* B*: Representation of the generation of a diffraction pattern.*

1.3.1.2 Crystals and the Unit Cell

A crystal can be described as the three-dimensional regular arrangement of molecules being in a low free-energy state. The lattice plane intersections occupied by arrays of atoms form the crystal and indicate this regularity (Fig. 1.7, A). The unit cell describes the bulk arrangement of atoms in a crystal and it is the smallest repetitive unit within this bulk (Fig. 1.7, B). Three repeating vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ with the angles between them, **α**, **β**, and **γ** define the unit cell. The x-axis of the crystal lattice goes along vector $\vec{a}$, the y-axis along $\vec{b}$, and the z-axis along $\vec{c}$. As it can be seen it in Fig. 1.7 B, the lattice intersections construct different planes. Depending on the plane orientation, either one, two, or all three lattice axes can be intersected by the lattice planes. Such a set of parallel lattice planes is determined by three indices, *h*, *k*, and *l*, the so-called Miller indices, which are used to label the corresponding reflections and describe the intensity of the diffracted beam. If a lattice set of planes is parallel to an axis, this index becomes zero, thus **100** describes an array of planes cutting the x-axis.

To calculate the electron density *p* of any position in space with coordinates x, y, and z, the so-called structure factors F_{hkl} are needed. They can be simply described as the sum of the diffraction contributions of all individual atoms in the unit cell. Thereby, the structure factor amplitude $|F_{hkl}|$ is proportional to the root of the intensity of the diffracted beam (eq. 3). Hence F_{hkl} describes mathematically the amplitude of the scattered X-ray beam and it can be seen as function of the electron density distribution in the unit cell.

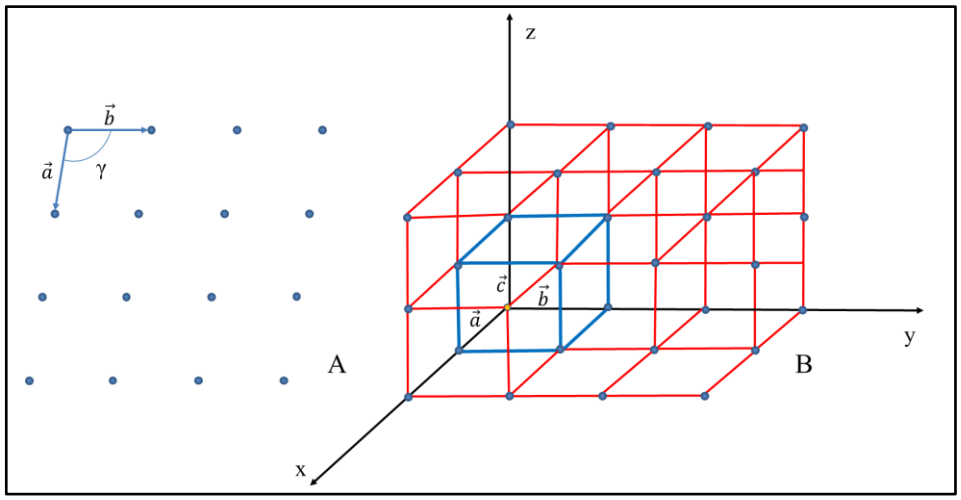


Fig. 1.7. A: *2D representation of the unit cell.* B: *A three-dimensional stack of unit cells describing a crystal lattice.*

$$|F_{hkl}| = \sqrt{I(hkl)} \qquad [3]$$

The entirety of all structure factors F_{hkl} can be translated by a Fourier transformation in an electron density distribution $p(xyz)$ in the unit cell (eq. 4), whereby V is the volume of the unit cell, x, y, and z are coordinates in the unit cell, and φ is the phase angle.

$$p(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|e^{[-2\pi i(hx+ky+lz)+i\varphi(hkl)]} \qquad [4]$$

However, although the structure factor amplitudes $|F_{hkl}|$ can be derived from the intensities $I(hkl)$ (eq. 3), it is not possible to extract the phase angles $\varphi(hkl)$ directly from the diffraction pattern. In crystallography, this is called the phase problem.

1.3.1.3 Friedel Pairs

As already mentioned in 1.3.1.1, the diffraction pattern of a crystal does not directly illustrate the crystal lattice but its reciprocal. Every reciprocal lattice has an inversion center (center of symmetry). Thus, every single reflection point on a certain *hkl* lattice plane produces a reflection point on the corresponding *-h-k-l* lattice plane. The intensities of both reflection spots are equal [I(*hkl*) = I(*-h-k-l*)], also referred to as Friedel's law ($F_{hkl} = F_{-h-k-l}$). F_{-h-k-l} is the mirror image of F_{hkl} leading to the fact that the Friedel pairs have opposite phases (Fig. 1.8, A). Under specific conditions of anomalous scattering the Friedel's law can be set false ($F_{hkl} \neq F_{-h-k-l}$) while solving the phase problem, e.g. by using heavy atoms or a specific X-ray wavelength (Fig. 1.8, B).

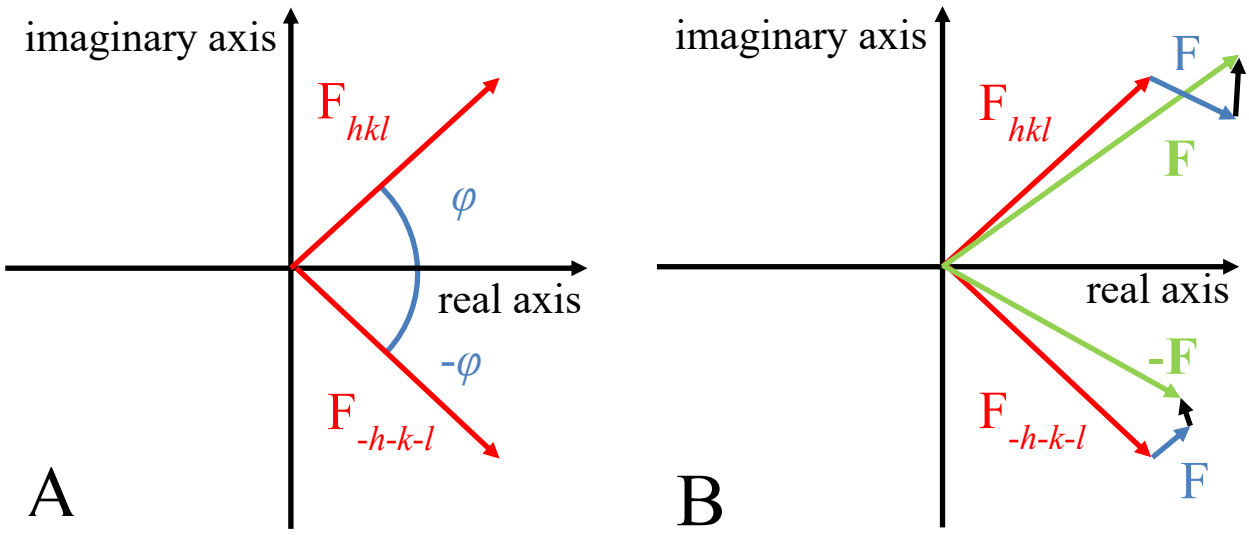


***Fig. 1.8**. Diagram for the structure factors of the reflections F_{hkl} and F_{-h-k-l}.*

1.3.1.4 Methods for Solving the Phase Problem

There are different ways to computationally solve the phase problem. The probably fastest one is that of the molecular replacement technique (MR). Thereby, the coordinates of a homologous protein are used to calculate the structure factor amplitudes F_{calc} and the phases φ_{calc}. In general, the procedure involves two steps. In the first, the spatial orientation of both molecules relative to one another is determined, followed by the second step, in which the translation in the unit cell is calculated, needed to overlay the now correctly oriented molecules onto one another. During the second step of MR, the so-called R-factor is calculated and used as target function in structure refinement. It results as a comparison between the calculated and observed structure factors (eq. 5).

$$R = \frac{\sum |F_{observed} - F_{calculated}|}{\sum |F_{observed}|} \qquad [5]$$

If no homologous protein model is available, the multiple isomorphous replacement (MIR) or multiple- and single-wavelength anomalous dispersion methods (MAD and SAD) can be used. For the MIR method, a dataset of the *apo* protein and datasets of at least two crystals with different heavy atoms have to be evaluated.[28] The heavy atoms cause a local shift of the electron density distribution. Thus, the dispersive difference between the structure factor amplitudes $|F_{hkl}|$ of the *apo* and derivative crystals are reduced to the presence of the heavy atom. The heavy atoms should not affect the dimensions of the unit cell of the derivative crystal compared to the *apo* crystal in order to make the method useful. The heavy atom component model can be calculated with the Patterson method, which is a Fourier transform with intensities as coefficients and phase angles equal to zero (eq. 6). The Patterson cell has coordinates *u*, *v*, and *w*, which correspond to the *x*, *y*, and *z* coordinates of a real cell.

$$P(uvw) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} |F_{hkl}|^2 \cos[2\pi(hu + kv + lw)] \qquad [6]$$

In contrast to MIR, in a MAD experiment only one crystal with introduced anomalous scatterer e.g. selenomethionine (SeMet) in place of methionine is needed. Data at several wavelengths near the absorption edge of the SeMet in the crystal are collected and along with the dispersive difference similar to MIR experiments also the anomalous difference is detected.

A big advantage of MAD compared to MIR is that non-isomorphism does not occur as all datasets are collected from one and the same crystal.
In a SAD experiment a single appropriate wavelength is used and only the anomalous differences are evaluated. For both, MAD and SAD experiments, the presence of a heavy atom in the crystal is not necessary but both require the presence of an anomalous scatterer and synchrotron radiation to tune the wavelength of the diffraction experiment.

1.3.2 Microscale Thermophoresis (MST)

Microscale thermophoresis is a rather new, immobilization-free method, describing the physical phenomenon, by which fluorescence-labeled molecules move in a temperature gradient (thermophoresis).[29] Thermophoresis is measured by monitoring the fluorescence distribution in a glass capillary. The temperature gradient is generated by an IR laser impulse and any induced diffusion processes are imaged by a CCD camera observing a fluorescence signal of the labeled protein molecules.
While in the initial phase the molecules are homogeneously distributed and a constant initial fluorescence ($F_{initial}$) of the labeled molecules is measured, a thermophoretic molecular movement toward cold regions is observed as soon as the IR-Laser is turned on, becoming evident in a fluorescence decrease (F_{hot}) (Fig. 1.9). The concentration of molecules in the heated region decreases until steady-state is reached. At a specific temperature increase ΔT, the molecular concentration ratio under this steady-state conditions is described by the Soret coefficient S_T (eq.7).

$$\frac{c_{hot}}{c_{cold}} = e^{(-S_T \cdot \Delta T)} \qquad [7]$$

When the IR-Laser is switched off, the molecules diffuse back, driven by properties such as mass, hydration shell, and charge, which can be used to detect binding. However, in an MST experiment a fluorescence signal resulting from the attached label at the protein is used to measure the molecular concentration changes, expressed in the normalized fluorescence F_{norm} (eq. 8). In recent studies the intrinsic fluorescence of tryptophan-containing proteins was used to monitor their thermophoresis, providing a label-free alternative to the standard approach.[30]

$$F_{norm} = \frac{F_{hot}}{F_{initial}} \quad [8]$$

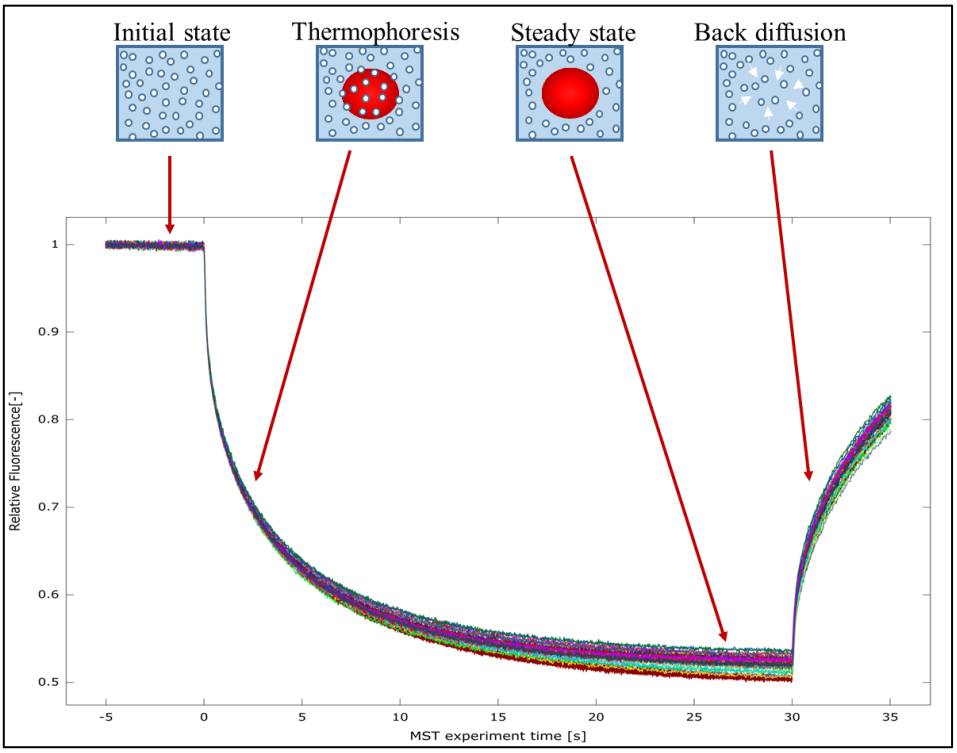


***Fig. 1.9**. In the initial state the molecules are homogeneously distributed. They diffuse toward colder regions upon heating with an IR laser until steady state is reached. When the IR laser is switched off the molecules diffuse back.*

In a usual MST experiment, the target of interest is labeled by a fluorescent probe which is additionally added to a dilution series of the ligand to be tested. The protein concentration should be lower or in the range of the dissociation constant K_d of the ligand of interest. Otherwise the inflection point of the binding curve would be shifted and inaccuracy in the determination of the K_d would occur.

Impure samples or aggregates being present in the solution can lead to high signal-to-noise ratios. Besides that, the quality of the capillaries is very important for a successful MST experiment. Some molecules tend to stick to the capillary surface leading to poor quality data, as only part of the molecules is mobile in the temperature gradient. In this case, covalently polymer-coated capillaries can be used instead.

1.4 Dynamic Combinatorial Chemistry

While the direct crystallographic screening of a fragment library still has a relative low throughput potential, the hit identification procedure can be speed up by a method called dynamic combinatorial chemistry (DCC).[31,32] Thereby, a dynamic combinatorial library (DCL) of ligands is generated by reactions between building blocks (Fig. 1.10). Under specific conditions the bonds between the building blocks are reversible and the system composed by multiple components reaches dynamic equilibrium. The product distribution is then dependent on the thermodynamic stability of the individual DCL members. The basic idea of this approach is the selection of the ligand with the highest affinity upon target protein addition under thermodynamic control.

A major advantage of this method compared to others is that a large number of DCL members is desired to speed up the identification of the best binder. The targeted synthesis of a single compound and its separation and purification from the large quantity of many side products in a conventional synthesis can be avoided, saving time and resources and the screening of multiple analogues is inherently performed in the presence of the target protein.

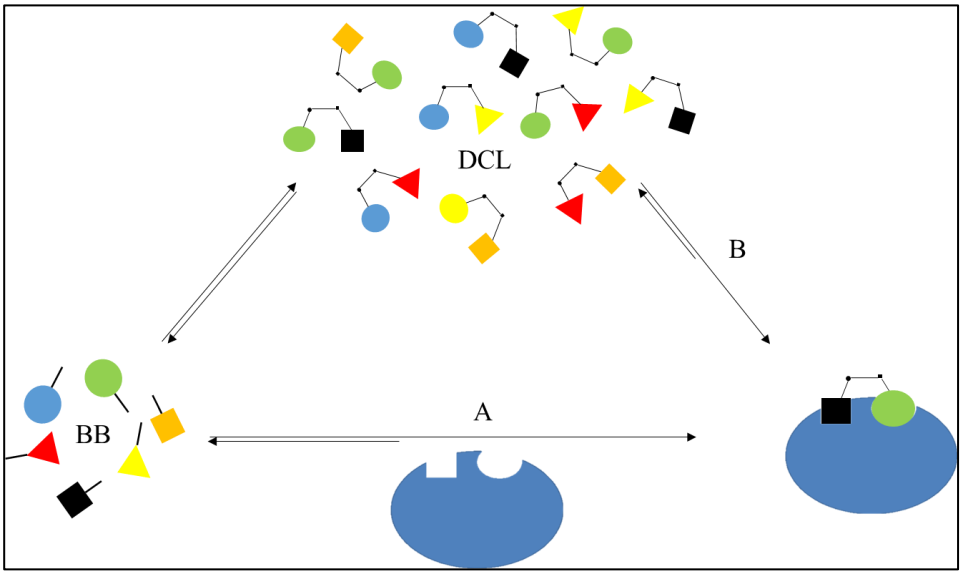


***Fig. 1.10**. Dynamic combinatorial chemistry. The target can be added either before (A) or after the system has reached equilibrium (B).*

On the other hand a high number of simultaneously present compounds with similar affinities can complicate the analysis since parameters such as compound solubility and chemical reactivity are strongly influenced and modulate the available concentrations under the test conditions. This will be discussed in chapter 6.

1.5 Aspartic Proteases

1.5.1 The Family of Aspartic Proteases

Aspartic proteases are a class of enzymes widespread in vertebrates, fungi, plants, and viruses. Their function is the hydrolytic cleavage of peptide substrates using a catalytically activated water molecule, which is placed and polarized as a nucleophile between two conserved aspartic acid residues. All family members have their optimum at acidic pH conditions and are characterized by the conserved position of the catalytic dyad. Remarkably, most of them are strongly inhibited by the natural product Pepstatin.[33,34]

The family of aspartic proteases includes a large number of subfamilies, as summarized in the MEROPS database.[35] Many family members are being intensively pursued as drug targets for therapeutic intervention in several diseases such as Alzheimer's disease (ß-site amyloid precursor protein cleaving enzyme 1, BACE1), hypertension (renin), mycosis (secreted aspartic proteases, SAP), malaria (plasmepsins), cancer (cathepsin D), and AIDS (human immunodeficiency virus type 1 protease (HIV1 protease)).

In general, the most abundant family members are divided into two subfamilies: pepsin-like (A1) and retroviral (A2) aspartic proteases.[36,37] The members of the A1 subfamily consist of one monomer (Fig. 1.11, left hand side). The active site, in the middle of which the catalytic aspartates can be found, is rather large and covered by the so-called flap region (Fig. 1.11, blue circle). This region is the most flexible part of all aspartic proteases as it has to open so that a substrate molecule or an inhibitor can enter the binding pocket and it closes again upon their binding. The A2 subfamily members are homodimers, which comprise two identical monomer units, each of which donates one aspartate residue for the composition of the catalytic dyad (Fig. 1.11, right hand side). Retroviral aspartic proteases are active only when both monomer units are present.

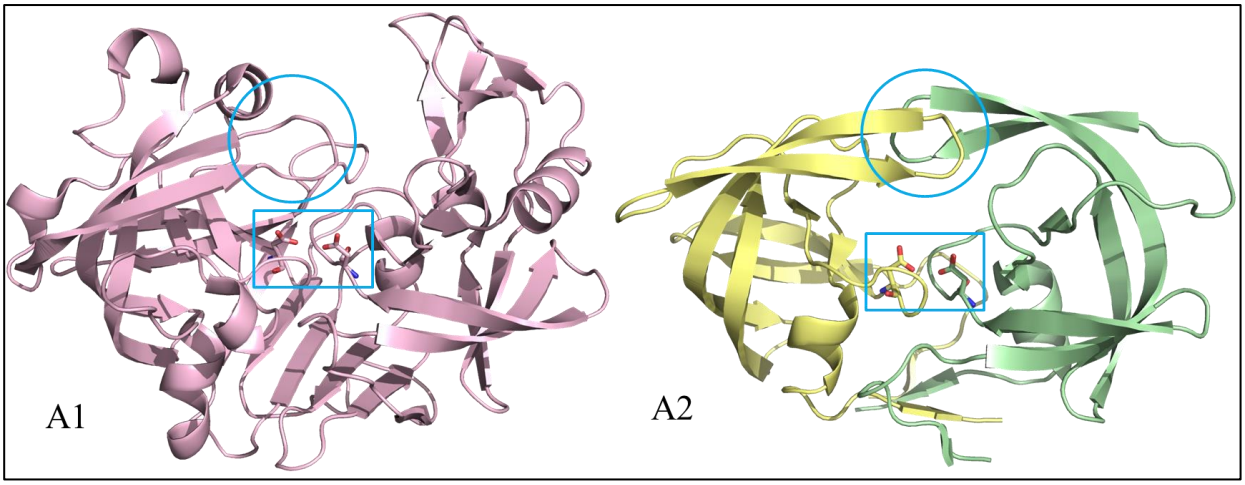


Fig. 1.11. *Cartoon presentation of the 3D structures of a pepsin-like (A1, PDB-code: 4LBT) and a retroviral (A2, PDB-code: 1SDT) aspartic protease. The flap region is highlighted with a blue circle and the catalytic dyad with a blue square.*

1.5.2 Functional Mechanism of Aspartic Proteases

As already mentioned all aspartic proteases have two catalytically active aspartic acid residues close to one another. Both residues have unusual pK_a values, one of ~ 2 and the other of ~ 5.[38] Thus, at acidic pH conditions one aspartate residue is protonated and the other one not. An important role for the cleavage of the substrate is taken over by the catalytic water molecule, which is positioned between the catalytic aspartates (Fig. 1.12). It gets strongly polarized by the deprotonated aspartate while the protonated one establishes an H-bond to the substrate's carbonyl group. This enables the nucleophilic attack at the carbonyl carbon atom performed by the catalytic water molecule. An intermediate tetrahedral geminal-diol transition state is only short-lived as it falls apart to the cleavage products, a carboxylic acid and a primary amine.

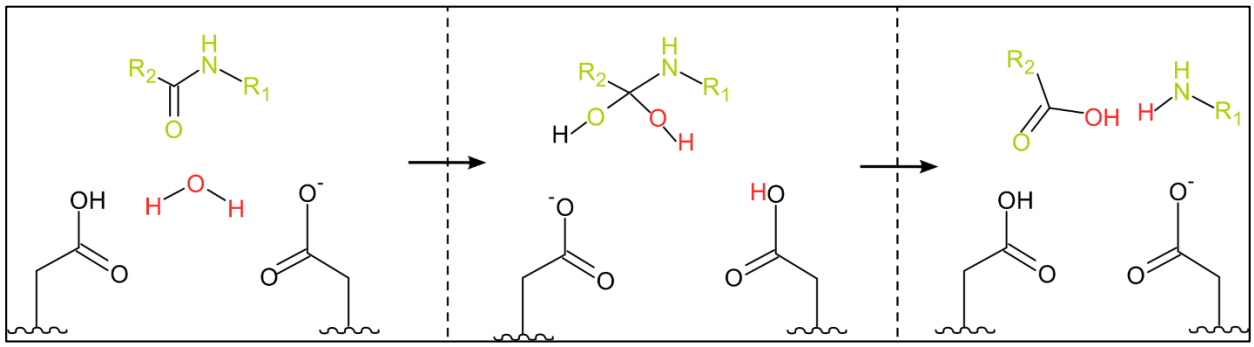


Fig. 1.12. *The catalytic mechanism of aspartic proteases.*

1.5.3 BACE1

BACE1, also known as membrane-associated aspartic protease 2 (memapsin 2), or aspartyl protease 2 (Asp 2), plays an important role in the early stage pathogenesis of Alzheimer's disease. It is a monomeric protein containing 456 amino acids, which cleaves full length ß-amyloid precursor protein (APP) into a soluble ectodomain fragment (sAPPß) and a C-terminal fragment (C99).[39] The further cleavage of C99 via γ-secretase releases an amyloid-ß protein (Aß) which accumulates in the brain and leads to the formation of neuritic plaques, ultimately resulting in dementia.[40,41] As biological studies proved that the Aß amount correlates with the degree of dementia in Alzheimer's disease patients, inhibiting the Aß formation at an early disease stage seems a promising therapeutic option.[41] As BACE1 is marginally expressed in the brain, challenges such as crossing the blood-brain barrier come along with the molecular size of peptidic inhibitors as well as selectivity features over other aspartic proteases making the research in this field very important.[42] The first developed BACE1 inhibitor is the peptide OM99-2 which was designed based upon a transition-state mimetic concept.[43] While this and many other peptidic inhibitors bind to the closed form of the enzyme, second generation compounds emerging from fragment-based screening techniques bind to the enzyme's open form and succeed to penetrate the nervous system (CNS). While no BACE1 inhibitor has been approved by the FDA so far, some second generation inhibitors have meanwhile reached clinical development stage involving AZD3839 from AstraZeneca,[44] LY2811376 from EliLilly,[45] (Fig. 1.13) and MK-8931 from Merck.[46]

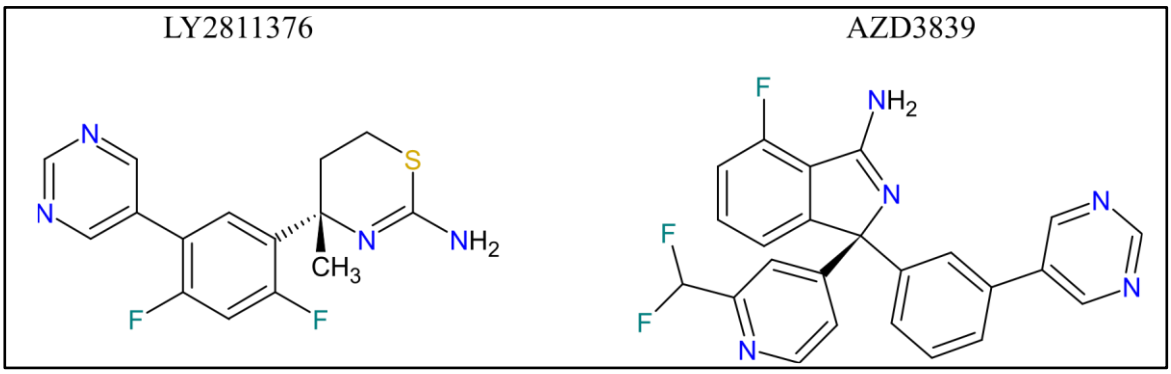


***Fig. 1.13**. 2D structures of clinical candidates targeting BACE1. The structure of MK-8931 is not known.*

1.5.4 Renin

Renin is a monomeric aspartic protease with 340 amino acids, which plays an important role in the blood pressure regulation. Produced in the kidneys, the enzyme cleaves its natural substrate angiotensinogen to the decapeptide angiotensin I. The latter is subsequently cleaved by the metalloprotease angiotensin converting enzyme (ACE) to the octapeptide angiotensin II which has a strong vasopressive effect on the blood vessels, upregulating the blood pressure.[47] After the successful marketing of ACE-inhibitors, many pharmaceutical companies concentrated on the search for selective renin inhibitors. A breakthrough succeeded at Ciba (now Novartis) with Aliskiren, the first and only marketed renin inhibitor so far, approved by the FDA in 2007.[48]

1.5.5 Secreted Aspartic Proteases

SAP proteins are secreted by Candida albicans, the main pathogen for candidiasis in humans.[49] The risks and consequences of the disease is particularly critical for immunocompromised patients, with diseases such as AIDS, cancer, or transplant ones, often leading to higher morbidity and mortality rates.[50] Candida albicans expresses ten different SAPs along with other virulence factors.[51] The best studied so far is the SAP2, a monomer consisting of 342 amino acids.
There are currently several commercially available topical and systemic agents for the treatment of candidiasis.[52] The gold standard is the intravenous use of amphotericin B given its large range of efficiency and low drug resistence.[53] Nonetheless, the steadily increasing mortality of candidiasis patients due to drug resistance makes the development of new antifungal agents urgent.[54]

1.5.6 Plasmepsins

Plasmepsins are produced by the protozoan parasite Plasmodium falciparum, which causes, along with Plasmodium vivax, Plasmodium ovale, and Plasmodium malariae, the most serious form of malaria in humans.[55] After a mosquito bite the parasite is directly transferred into the blood and the infection occurs within 10 – 14 days. The plasmepsins digest hemoglobin, which breakdown products are used by the parasite for nutrition.

For the treatment of uncomplicated malaria an artemisinin-based combination therapy (ACT) is recommended by the world health organization (WHO).[56] The biggest problem in the antimalarial therapy is the drug resistance emerging against all antimalarial drugs.[57]

1.5.7 HIV1 Protease

The HIV1 protease belongs to the A2 subfamily. The protease represents a homodimer consisting of two monomer units of 99 amino acids each.[58] It is involved in the pathogenesis of the acquired immunodeficiency syndrome (AIDS) which was first described in 1981. According to the United Nations Programme on HIV/AIDS (UNAIDS) there are currently 36.7 million people living with HIV. The virus causing the disease is transmitted through body fluids encountering damaged tissues or mucous membrane or by the direct transfer into the bloodstream. The HIV1 protease is required for the viral replication. Thus, the enzyme inhibition reduces the reproduction of the virus. The first HIV1 protease inhibitor approved by the FDA in 1995 is Saquinavir followed by the market launch of currently eight further inhibitors. Due to the resistance development of the HI virus, compliance issues in therapies caused by serious side effects, and the high therapy costs make the development of new antiviral drugs urgent.

1.5.8 Endothiapepsin

Endothiapepsin (EP) is a monomer from the A1 subfamily of aspartic proteases, comprising 330 amino acids. It is produced by the fungus Cryphonectria parasitica which is the main cause of chestnut blight.[59] The fungus is not dangerous to humans.

EP is used in the industry for the production of cheese due to its lactic coagulation properties.[60] It is commercially available under the name Suparen®. However, for research purposes EP should be first purified from Suparen®. Furthermore, there is no expression system available for the protease, which makes mutations within the backbone at present impossible.

In the field of drug discovery, EP has been used as a model protein for the development of renin and BACE1 inhibitors since the early eighties.[61,62] As EP was the main target investigated in this work, the quality of the binding site and key pocket residues will be described in the following. EP comprises nine binding pockets in its binding site (Fig. 1.14). The S1’ pocket, surrounded by Ile300, Ile302, and Ile304, is the most hydrophobic one.

The S2’ pocket is less hydrophobic as it is surrounded by Ile77, Leu133, and Phe194. The adjacent S3’ pocket is above the S1’ and S2’ pockets. It exhibits the polar residues Ser78, Tyr79, and Gly80 and it thus can be classified as hydrophilic. Hydrophilic is also the S2 pocket, which is surrounded by Thr222, Thr223, and Tyr226. The S3 pocket, surrounded by Ile10, Ala16 and Ile122 is partly covered by the flap region. It is adjacent to the S1 pocket, which is surrounded by Tyr79, Phe116 and Leu125. Both, the S3 and S1 pockets are thus classified as amphiphilic. The solvent exposure of the S4 and S5 pockets renders these with hydrophilic character. The S6 pocket is also exposed to the solvent. However, it is surrounded by the hydrophobic amino acids Leu13, Phe280, Ile283 and Phe291.

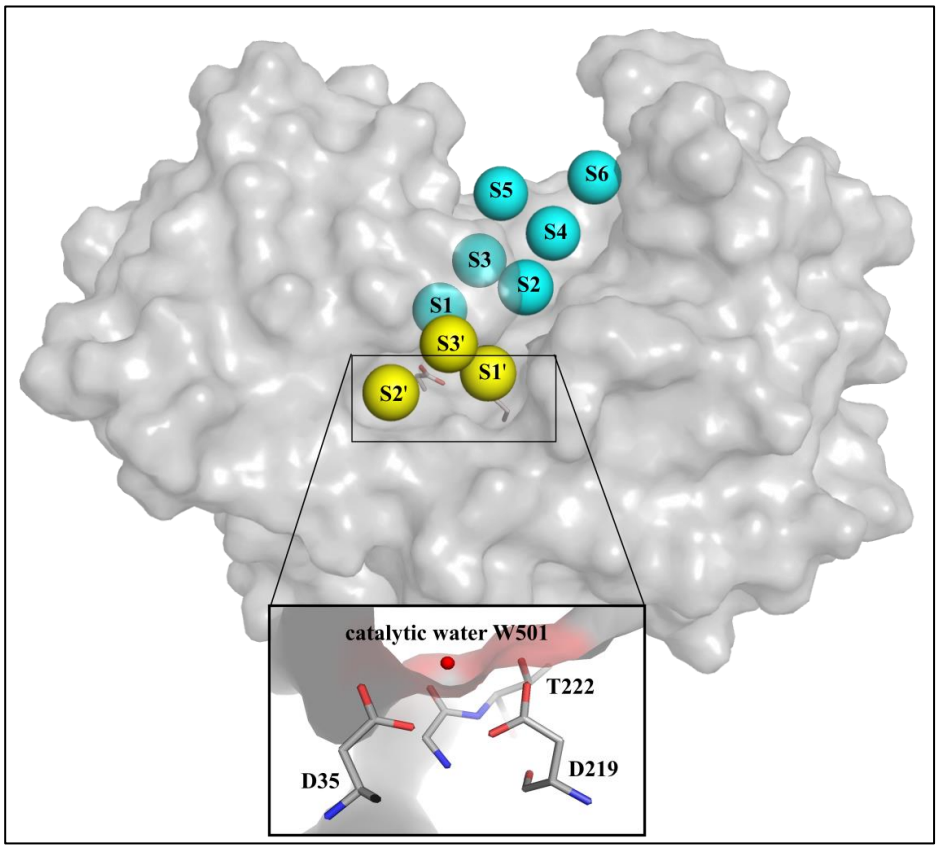


Fig. 1.14. *Surface representation of EP with its binding pockets depicted as spheres.*

Although the sequence identity with other aspartic proteases is around 30%, the use of EP in the basic research field for determining the interaction pattern of small molecules bound to the target is still a common procedure as it allows one to achieve high quality data with up to 1Å resolution, which is hardly achievable without spending significant effort for any of the other relevant drug targets.[63] However, structurally identical regions are observed as well as differences, which can be used to circumvent putative selectivity issues.

For example, by comparing the structure of EP with those of BACE1, renin, cathepsin D, and SAP2, conserved regions beyond the catalytic dyad can be observed.[26] All proteases share a tyrosine residue, which resides on the flap region (Fig. 1.15, A). Furthermore, also the position of a disulfide bridge (Cys255 and Cys291 in EP) is found conserved in all proteases (Fig. 1.15, B). However, in a recent study of EP we found hot spots of binding such as Asp81 neighboring the conserved Tyr residue but also Asp119 further away from the catalytic dyad and Phe291 in the large S6 pocket. Apparently, all three residues can be found at the same position only in SAP2.

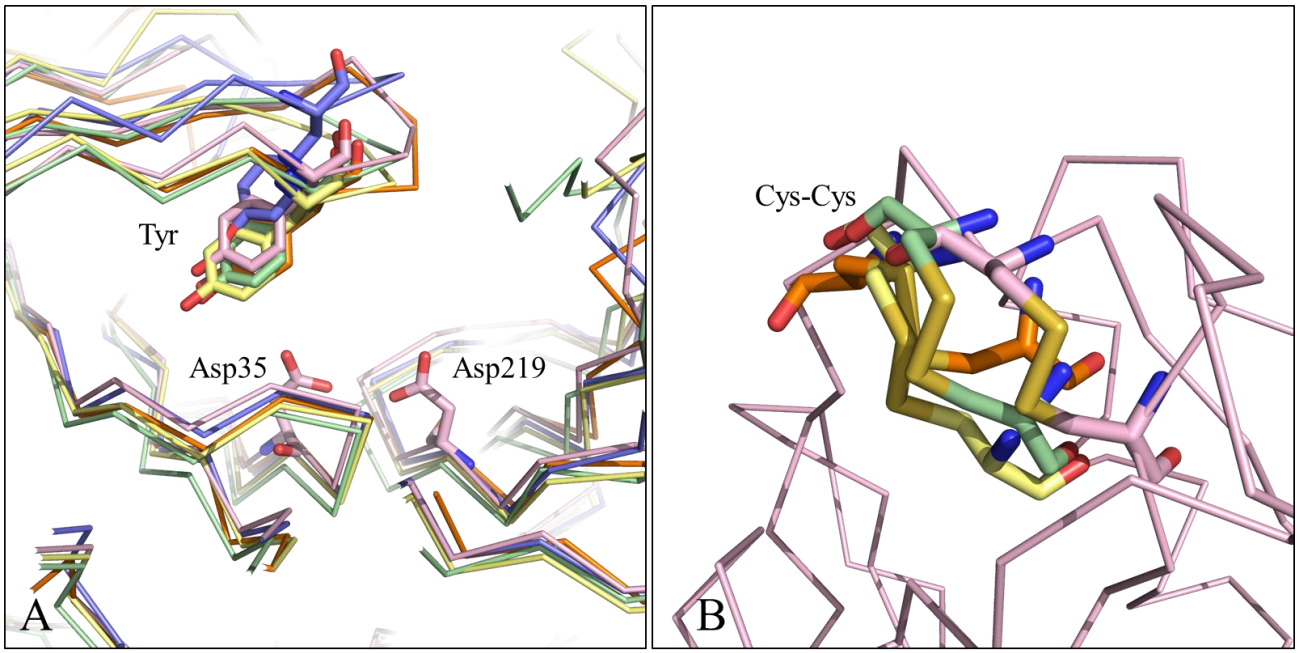


***Fig. 1.15**. Five aspartic proteases are superimposed (EP in pink (PDB-code: 4LBT), BACE1 in purple (PDB-code: 5ENM), renin in green (PDB-code: 4RYC), cathepsin D in yellow (PDB-code: 4OC6), and SAP2 in orange (PDB-code: 1EAG)). A: All proteases are shown as ribbon apart from the catalytic dyad of EP and the overlapping Tyr residues of the flap region. B: The conserved disulfide bridges are depicted as sticks and the EP backbone is shown as ribbon. The disulfide bond of BACE1 was omitted for more clarity as the bond was reduced in that particular BACE1 structure.*

1.6 Preliminary Work and Aims of the Thesis

In the framework of this thesis the identification and characterization of novel efficient binders with low molecular weight for the aspartic protease Endothiapepsin (EP) has been performed. For this purpose X-ray crystallography was consulted as primary screening method and its hit identification potential compared to several biochemical and biophysical screening methods, performed in cooperation with industry.

The work was initiated in the past by colleagues from the group of Prof. Klebe, who designed a fragment library for general purposes. Initially, all fragment library entries were screened against EP via a fluorescence-based assay and obtained hits have been analyzed crystallographically using cocktails of two. The biochemical assay resulted in 17% hit rate, while the X-ray analysis in only 3%.

The work in this thesis started at this point. It comprised the following aims:

1. The setup and performance of a microscale thermophoresis (MST) assay with the in-house fragment library on EP. In addition, the method's capability of identifying fragment hits was compared to other screening techniques.
2. Due to the unsatisfactory low overlap of hit lists between the analyzed screening techniques, X-ray crystallography was consulted in addition. Thereby, single soaking experiments with the in-house library as well as the additional data collection and analysis of overall more than 450 datasets were performed. The main aim of the crystallographic analysis, used as primary screening technique, was the comparison of the yielded hit rates with those of the initially performed assays as well as the determination of binding modes and interaction patterns of the fragments with EP.
3. The analysis and comparison of cocktails with single soaking experiments. It was aimed to demonstrate that the use of cocktails to accelerate hit identification often leads to loss of hits.
4. In a separate project, dynamic combinatorial chemistry was investigated as a promising alternative for the discovery of potent hits involving the target protein in the hit selection step.

1.7 References

[1] Kalyaanamoorthy, S.; Chen, Y.-P. P. Structure-based drug design to augment hit discovery. *Drug Discovery Today* **2011**, *16*, 831–839.

[2] Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **2003**, *10*, 787–797.

[3] Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649–663.

[4] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

[5] Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discovery* **2010**, *9*, 203–214.

[6] DiMasi, J. A.; Grabowski, H. G. The cost of biopharmaceutical R&D: Is biotech different? *Manage. Decis. Econ.* **2007**, *28*, 469–479.

[7] Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discovery* **2011**, *10*, 188–195.

[8] Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* **1996**, *16*, 3–50.

[9] Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci.* **1999**, *96*, 9997–10002.

[10] Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: A useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430–431.

[11] Doak, B. C.; Norton, R. S.; Scanlon, M. J. The ways and means of fragment-based drug design. *Pharmacol. Ther.* **2016.** DOI: 10.1016/j.pharmthera.2016.07.003.

[12] Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discovery* **2016**, *15*, 605–619.

[13] Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004**, *3*, 660–672.

[14] Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 4046–4050.

[15] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discovery Today* **2003**, *8*, 876–877.

[16] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26.

[17] Jhoti, H.; Williams, G.; Rees, D. C.; Murray, C. W. The 'rule of three' for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discovery* **2013**, *12*, 644–645.

[18] Hall, R. J.; Mortenson, P. N.; Murray, C. W. Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* **2014**, *116*, 82–91.

[19] Hann, M. M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.

[20] Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. The Design of Leadlike Combinatorial Libraries. *Angew. Chem., Int. Ed.* **1999**, *38*, 3743–3748.

[21] Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K.; Sproesser, K.; Li, L.; Smalley, K. S. M.; Fong, D.; Zhu, Y.-L.; Marimuthu, A.; Nguyen, H.; Lam, B.; Liu, J.; Cheung, I.; Rice, J.; Suzuki, Y.; Luu, C.; Settachatgul, C.; Shellooe, R.; Cantwell, J.; Kim, S.-H.; Schlessinger, J.; Zhang, K. Y. J.; West, B. L.; Powell, B.; Habets, G.; Zhang, C.; Ibrahim, P. N.; Hirth, P.; Artis, D. R.; Herlyn, M.; Bollag, G. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3041–3046.

[22] Souers, A. J.; Leverson, J. D.; Boghaert, E. R.; Ackler, S. L.; Catron, N. D.; Chen, J.; Dayton, B. D.; Ding, H.; Enschede, S. H.; Fairbrother, W. J.; Huang, D. C. S.; Hymowitz, S. G.; Jin, S.; Khaw, S. L.; Kovar, P. J.; Lam, L. T.; Lee, J.; Maecker, H. L.; Marsh, K. C.; Mason, K. D.; Mitten, M. J.; Nimmer, P. M.; Oleksijew, A.; Park, C. H.; Park, C.-M.; Phillips, D. C.; Roberts, A. W.; Sampath, D.; Seymour, J. F.; Smith, M. L.; Sullivan, G. M.; Tahir, S. K.; Tse, C.; Wendt, M. D.; Xiao, Y.; Xue, J. C.; Zhang, H.; Humerickhouse, R. A.; Rosenberg, S. H.; Elmore, S. W. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **2013**, *19*, 202–208.

[23] Schiebel, J.; Radeva, N.; Koster, H.; Metz, A.; Krotzky, T.; Kuhnert, M.; Diederich, W. E.; Heine, A.; Neumann, L.; Atmanene, C.; Roecklin, D.; Vivat-Hannah, V.; Renaud, J.-P.; Meinecke, R.; Schlinck, N.; Sitte, A.; Popp, F.; Zeeb, M.; Klebe, G. One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* **2015**, *10*, 1511–1521.

[24] Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J. Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat. Rev. Drug Discovery* **2016**, *15*, 533–550.

[25] Cox, D.; Brennan, M.; Moran, N. Integrins as therapeutic targets: lessons and opportunities. *Nat. Rev. Drug Discovery* **2010**, *9*, 804–820.

[26] Radeva, N.; Krimmer, S. G.; Stieler, M.; Fu, K.; Wang, X.; Ehrmann, F. R.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Schiebel, J.; Heine, A.; Klebe, G. Experimental Active-Site Mapping by Fragments: Hot Spots Remote from the Catalytic Center of Endothiapepsin. *J. Med. Chem.* **2016**, *59*, 7561–7575.

[27] Schiebel, J.; Krimmer, S. G.; Rower, K.; Knorlein, A.; Wang, X.; Park, A. Y.; Stieler, M.; Ehrmann, F. R.; Fu, K.; Radeva, N.; Krug, M.; Huschmann, F. U.; Glockner, S.; Weiss, M. S.; Mueller, U.; Klebe, G.; Heine, A. High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* **2016**, *24*, 1398–1409.

[28] Green, D. W.; Ingram, V. M.; Perutz, M. F. The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method. *Proc. R. Soc. A* **1954**, *225*, 287–307.

[29] Baaske, P.; Wienken, C. J.; Reineck, P.; Duhr, S.; Braun, D. Optical thermophoresis for quantifying the buffer dependence of aptamer binding. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 2238–2241.

[30] Seidel, S. A. I.; Wienken, C. J.; Geissler, S.; Jerabek-Willemsen, M.; Duhr, S.; Reiter, A.; Trauner, D.; Braun, D.; Baaske, P. Label-free microscale thermophoresis discriminates sites and affinity of protein-ligand binding. *Angew. Chem., Int. Ed. Engl.* **2012**, *51*, 10656–10659.

[31] Lehn, J.-M. From supramolecular chemistry towards constitutional dynamic chemistry and adaptive chemistry. *Chem. Soc. Rev.* **2007**, *36*, 151–160.

[32] Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M.; Otto, S. Dynamic combinatorial chemistry. *Chem. Rev.* **2006**, *106*, 3652–3711.

[33] Seelmeier, S.; Schmidt, H.; Turk, V.; Helm, K. von der. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 6612–6616.

[34] Rich, D. H.; Bernatowicz, M. S.; Agarwal, N. S.; Kawai, M.; Salituro, F. G.; Schmidt, P. G. Inhibition of aspartic proteases by pepstatin and 3-methylstatine derivatives of pepstatin. Evidence for collected-substrate enzyme inhibition. *Biochem.* **1985**, *24*, 3165–3173.

[35] Rawlings, N. D.; Barrett, A. J.; Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Res.* **2010**, *38*, 227–233.

[36] Dunn, B. M. Overview of pepsin-like aspartic peptidases. *Curr. Protoc. Protein Sci.* **2001**, *Chapter 21*, Unit 21.3.

[37] Szecsi, P. B. The aspartic proteases. *Scand. J. Clin. Lab. Invest., Suppl.* **1992**, *210*, 5–22.

[38] Pearl, L.; Blundell, T. The active site of aspartic proteinases. *FEBS Letters* **1984**, *174*, 96–101.

[39] Selkoe, D. J. Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.* **2001**, *81*, 741–766.

[40] Krishnaswamy, S.; Verdile, G.; Groth, D.; Kanyenda, L.; Martins, R. N. The structure and function of Alzheimer's gamma secretase enzyme complex. *Crit. Rev. Clin. Lab. Sci.* **2009**, *46*, 282–301.

[41] Hardy, J.; Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **2002**, *297*, 353–356.

[42] Broccatelli, F.; Larregieu, C. A.; Cruciani, G.; Oprea, T. I.; Benet, L. Z. Improving the prediction of the brain disposition for orally administered drugs using BDDCS. *Adv. Drug Delivery Rev.* **2012**, *64*, 95–109.

[43] Shimizu, H.; Tosaki, A.; Kaneko, K.; Hisano, T.; Sakurai, T.; Nukina, N. Crystal structure of an active form of BACE1, an enzyme responsible for amyloid beta protein production. *Mol. Cell. Biol.* **2008**, *28*, 3663–3671.

[44] Jeppsson, F.; Eketjall, S.; Janson, J.; Karlstrom, S.; Gustavsson, S.; Olsson, L.-L.; Radesater, A.-C.; Ploeger, B.; Cebers, G.; Kolmodin, K.; Swahn, B.-M.; Berg, S. von; Bueters, T.; Falting, J. Discovery of AZD3839, a potent and selective BACE1 inhibitor clinical candidate for the treatment of Alzheimer disease. *J. Biol. Chem.* **2012**, *287*, 41245–41257.

[45] May, P. C.; Dean, R. A.; Lowe, S. L.; Martenyi, F.; Sheehan, S. M.; Boggs, L. N.; Monk, S. A.; Mathes, B. M.; Mergott, D. J.; Watson, B. M.; Stout, S. L.; Timm, D. E.; Smith Labell, E.; Gonzales, C. R.; Nakano, M.; Jhee, S. S.; Yen, M.; Ereshefsky, L.; Lindstrom, T. D.; Calligaro, D. O.; Cocke, P. J.; Greg Hall, D.; Friedrich, S.; Citron, M.; Audia, J. E. Robust central reduction of amyloid-beta in humans with an orally available, non-peptidic beta-secretase inhibitor. *J. Neurosci.* **2011**, *31*, 16507–16516.

[46] Stamford, A.; Strickland, C. Inhibitors of BACE for treating Alzheimer's disease: a fragment-based drug discovery story. *Curr. Opin. Chem. Biol.* **2013**, *17*, 320–328.

[47] Vecsei, P.; Hackenthal, E.; Ganten, D. The renin-angiotensin-aldosterone system. *Klin. Wochenschr.* **1978**, *56*, 5–21.

[48] Sanoski, C. A. Aliskiren: an oral direct renin inhibitor for the treatment of hypertension. *Pharmacother.* **2009**, *29*, 193–212.

[49] Gropp, K.; Schild, L.; Schindler, S.; Hube, B.; Zipfel, P. F.; Skerka, C. The yeast Candida albicans evades human complement attack by secretion of aspartic proteases. *Mol. Immunol.* **2009**, *47*, 465–475.

[50] Borg-von Zepelin, M.; Kunz, L.; Ruchel, R.; Reichard, U.; Weig, M.; Gross, U. Epidemiology and antifungal susceptibilities of Candida spp. to six antifungal agents: results from a surveillance study on fungaemia in Germany from July 2004 to August 2005. *J. Antimicrob. Chemother.* **2007**, *60*, 424–428.

[51] Schaller, M.; Borelli, C.; Korting, H. C.; Hube, B. Hydrolytic enzymes as virulence factors of Candida albicans. *Mycoses* **2005**, *48*, 365–377.

[52] Gupta, A. K.; Sauder, D. N.; Shear, N. H. Antifungal agents: an overview. Part I. *J. Am. Acad. Dermatol.* **1994**, *30*, 677–698.

[53] Moen, M. D.; Lyseng-Williamson, K. A.; Scott, L. J. Liposomal amphotericin B: a review of its use as empirical therapy in febrile neutropenia and in the treatment of invasive fungal infections. *Drugs* **2009**, *69*, 361–392.

[54] Yapar, N. Epidemiology and risk factors for invasive candidiasis. *Ther. Clin. Risk Manage.* **2014**, *10*, 95–105.

[55] Trager, W.; Jensen, J. Human malaria parasites in continuous culture. *Science* **1976**, *193*, 673–675.

[56] *Guidelines for the treatment of malaria*, Third edition; W. H. O.: Geneva, 2015.

[57] Sinha, S.; Medhi, B.; Sehgal, R. Challenges of drug-resistant malaria. *Parasite* **2014**, *21*, 61–76.

[58] Davies, D. R. The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189–215.

[59] Barakat, A.; DiLoreto, D. S.; Zhang, Y.; Smith, C.; Baier, K.; Powell, W. A.; Wheeler, N.; Sederoff, R.; Carlson, J. E. Comparison of the transcriptomes of American chestnut (Castanea dentata) and Chinese chestnut (Castanea mollissima) in response to the chestnut blight infection. *BMC Plant Biol.* **2009**, *9*, 51.

[60] Rawlings, N. D.; Salvesen, G. *Handbook of proteolytic enzymes*, Third edition; Elsevier Academic: Amsterdam, London, 2013.

[61] Geschwindner, S.; Olsson, L.-L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; Beer, T. de; Folmer, R. H. A. Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J. Med. Chem.* **2007**, *50*, 5903–5911.

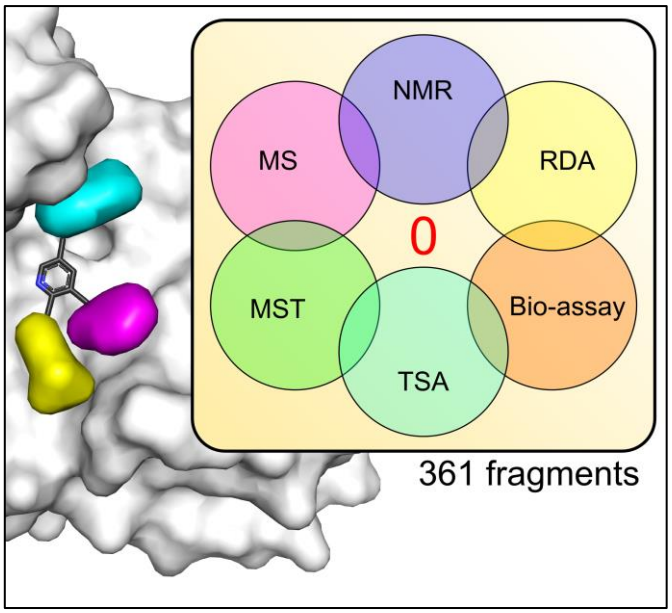
[62] Cooper, J.; Quail, W.; Frazao, C.; Foundling, S. I.; Blundell, T. L.; Humblet, C.; Lunney, E. A.; Lowther, W. T.; Dunn, B. M. X-ray crystallographic analysis of inhibition of endothiapepsin by cyclohexyl renin inhibitors. *Biochemistry* **1992**, *31*, 8142–8150.

[63] Radeva, N.; Schiebel, J.; Wang, X.; Krimmer, S. G.; Fu, K.; Stieler, M.; Ehrmann, F. R.; Metz, A.; Rickmeyer, T.; Betz, M.; Winquist, J.; Park, A. Y.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Heine, A.; Klebe, G. Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures:Versatile Warheads to Address a Catalytic Dyad. *J. Med. Chem.* **2016**, *59*, 9743–9759.

2. One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists

2.1 Introductory Remarks

The following chapter has been published in *ChemMedChem* in 2015. The author of the thesis developed and performed the MST assay and analysis in collaboration with Nina Schlinck and Astrid Sitte (NanoTemper) as well as Franziska Popp and Markus Zeeb (Boehringer Ingelheim). Furthermore, Helene Köster performed the biochemical assay, the NMR measurements were developed and performed by Robert Meinecke and Markus Zeeb (Boehringer Ingelheim), Maren Kuhnert and Wibke E. Diederich contributed the TSA measurements, the ESI-MS analyses were performed by Cedric Atmanene, Dominique Roecklin, Valérie Vivat-Hannah, and Jean-Paul Renaud (NovAliX), the RDA say was performed by Lars Neumann (Proteros), the assay-data correlation and the paper drafting were performed by Johannes Schiebel.



*Reprinted with permission from *ChemMedChem* **2015**, *10*, 1511–1521. License number: 3995361210259.

2.2 Abstract

Fragment-based lead discovery is gaining momentum in drug development. Typically, a hierarchical cascade of several screening techniques is consulted to identify fragment hits which are then analyzed by crystallography. Because crystal structures with bound fragments are essential for the subsequent hit-to-lead-to-drug optimization, the screening process should distinguish reliably between binders and non-binders. We therefore investigated whether different screening methods would reveal similar collections of putative binders. First we used a biochemical assay to identify fragments that bind to endothiapepsin, a surrogate for disease relevant aspartic proteases. In a comprehensive screening approach, we then evaluated our 361-entry library by using a reporter-displacement assay, saturation-transfer difference NMR, native mass spectrometry, thermophoresis, and a thermal shift assay. While the combined results of these screening methods retrieve 10 of the 11 crystal structures originally predicted by the biochemical assay, the mutual overlap of individual hit lists is surprisingly low, highlighting that each technique operates on different biophysical principles and conditions.

2.3 Introduction

Fragment-based lead discovery (FBLD) has matured over the last decade to a reliable and efficient method, generating promising leads for subsequent medicinal chemistry optimization.[1,2] Stimulated by the low productivity and huge costs of high-throughput screening, compound acquisition, management, and handling, FBLD has emerged as a technique that can address a large chemical space by using smaller compounds.[3–5] Nevertheless, even though the initial fragment hits are rather weak milli- to micromolar binders, their optimization can be carried out quite efficiently, particularly when supported by a structure-based approach.[6,7]

The field was initiated by the screening of solvent molecules as molecular probes to map protein binding sites by crystallography[8] or by tailored searches for molecular “needles” that bind the hot spots of binding pockets.[9] NMR emerged in the 1990s as the method of choice to detect weak binding fragments, to reveal initial structural insight into their binding modes, and to guide initial optimization steps.[10] The major disadvantages of this “SAR by NMR” approach are the need for isotope-labeled protein, full assignment of the resonances, and its limitation to rather small proteins. Therefore, over time an increasing number of biochemical and biophysical methods have been developed and optimized regarding their sensitivity to detect fragment binding.[11,12]

Apart from the protein-observed NMR technique that requires isotope labeling, less elaborate one-dimensional 1H NMR spectroscopic methods have been developed, such as saturation-transfer difference (STD) NMR[13] and water-LOGSY[14] experiments. The latter ligand-observed NMR methods exploit the transfer of magnetization from protein (or binding site water molecules) to a transiently bound fragment. Relaxation-edited NMR experiments evaluate differences in the relaxation properties of small and large molecules and use this information for fragment discovery. Techniques based on ^{19}F NMR resonance detecting the replacement of a fluorinated spy ligand can also be applied to detect fragment binding.[15,16]

Functional biochemical assays can sometimes be optimized in a way that they still work under high-concentration conditions, allowing the detection of inhibition by weak binding fragments.[17] The displacement of radioactively labeled probe molecules is an alternative to detect binding.[18] This method is very sensitive; however, it requires a labeled reference ligand. Similarly, a reporter displacement assay (RDA) can be used.[19,20] This approach uses a reporter probe molecule labeled with an optically sensitive group. Proximity between reporter and protein results in the emission of an optical signal, and fragments capable of displacing the reporter ligand will cause a loss of signal.

A popular, simple, and fast assay is based on the fact that the temperature-dependent unfolding of a protein is shifted toward higher temperature once a ligand stabilizes a macromolecule. In this thermal shift assay (TSA), protein unfolding is monitored using the fluorescence of a dye, which is enhanced once hydrophobic portions exposed by the unfolding protein provide a platform on which the dye molecules start to assemble.[21–23] In recent years, native electrospray ionization mass spectrometry (ESI-MS) has emerged as a promising label-free technique for medium-throughput screening of small-molecule libraries against protein targets.[24–26] In native MS, compound binding stoichiometry and relative binding affinity information can be obtained from accurate mass measurements and relative signal intensities of the free protein and related protein–ligand complex, respectively. Alternative methods require the probe molecules to be immobilized on a sensor surface, and changes in the optical properties of the surface material are monitored, for example, the refractive index close to a thin metal film using an evanescent field.[12,27,28] These surface plasmon resonance (SPR) techniques usually work with an immobilization of the protein and detect fragment binding. However, also the reverse concept has been exploited by immobilizing ligands as exposed chemical microarrays.[29] As a medium-throughput method, isothermal titration calorimetry (ITC) can also be used to measure the heat signal produced upon

complex formation. However, it requires sophisticated titration protocols and large amounts of protein material.[11,30,31] As a rather new technology, microscale thermophoresis (MST) has emerged, and its scope has been expanded toward fragment discovery.[32] The method records the motion of fluorescence-labeled molecules in a microscopic temperature gradient. Recent developments also exploit the fluorescence of tryptophan to trace molecular motions. Small changes in the hydration shell of the observed macromolecule, owing to the accommodation of a small molecule fragment, modulate the thermophoretic movement which can be used to detect binding.

A major disadvantage all these approaches have in common is that they do not provide detailed information about the binding geometry of the discovered fragment hits. Nonetheless, the fundamental consensus that structural information is crucial to an FBLD project to guide the subsequent rational potency enhancement in fragment optimization emerged within the medicinal chemistry community. At this point, crystal structure analysis is the method of choice.[11,33,34] However, as crystallography is still a rather elaborate technique, which requires a robust crystallization system producing easily soakable and well-diffracting crystals, it is usually applied as a follow-up method only considering a short list of the most promising fragment hits detected by the less demanding "pre"-screening approaches (the term *prescreening* is used in the following to highlight that those techniques are usually applied to select fragments for the subsequent crystallographic screening).[11] This kind of stepwise screening cascade is based on the important assumption that the coverage of the prescreening methods with respect to possible binders is quite exhaustive and reliable. In case they retrieve false positive hits, the follow-up crystallographic screening will disclose such inadequacies even though it will then become less efficient. The situation becomes more cumbersome if prescreening fails owing to false negatives, as they will not be considered in a subsequent crystallographic study and, thus, decrease the number of crystal structures that are the basis for subsequent medicinal chemistry efforts.

We recently presented a newly compiled fragment library,[35] which had been assembled by rational criteria oriented toward, but not strictly adhering to, the rule of three.[36] This original compilation contains 364 entries including three accidentally overlooked duplicates, leading to 361 non-redundant fragments in the final library. The design of the library and the compounds selected for fragment screening is crucial.[37,38] Apart from aiming at high solubility, we designed our library to contain fragments which on average possess a larger portion of Lipinski acceptor functionalities. The quality of our library was tested using a biochemical assay on

endothiapepsin, an enzyme frequently used as relevant surrogate for the large class of pepsin-like aspartic proteases.[35] This family of enzymes comprises a number of important disease targets such as malaria (plasmepsins), fungal infections (secreted aspartic proteinases), Alzheimer's disease (b-secretase), and hypertension (renin). Recently, we demonstrated that this protein is a real challenge in drug discovery due to its large binding pocket. In particular, a variety of distinct binding modes could be detected for a congeneric inhibitor series that only slightly varied at different positions of its parent scaffold.[39] The crystallographic fragment screening of all 55 hits from the biochemical assay led to 11 fragment complexes with endothiapepsin. These different fragments cover the binding cleft of the target enzyme quite broadly. Remarkably, consulting the rule of three as a putative selection criterion for the library design, we would have only discovered four of the eleven hits. The additional seven fragments would have been missed, as they all violate the rule of three.

As mentioned, we followed a cascade approach using crystallography to validate and structurally characterize the hits suggested by prescreening with the high-concentration biochemical assay based on the cleavage of a fluorescent peptide. The question arises whether an alternative prescreening method would have indicated the same or a distinct sample of promising hits to be followed up as promising candidates in a subsequent crystallographic study. To the best of our knowledge there is no comprehensive study available which tested a full fragment library under similar conditions using a series of state-of-the-art fragment screening techniques. We therefore decided to consult, apart from the already reported biochemical assay, five other prescreening techniques to enumerate our 361-entry library for putative binders. A thermal shift assay (TSA), STD-NMR, a reporter-displacement assay (RDA), native nano-ESI mass spectrometry (ESI-MS), and microscale thermophoresis (MST) were used to screen our library for promising fragment hits. The results of these screening campaigns are compared with the crystallographic hits, which were obtained based on the analysis of all fragments significantly inhibiting endothiapepsin in the biochemical assay.

2.4 Results and Discussion

In our previously reported fragment screening campaign we focused entirely on the results of a biochemical high-concentration prescreening assay. Only fragments inhibiting endothiapepsin to at least 40% were subjected to a subsequent crystallographic study to further characterize these

most promising hits.[35] Out of the 55 fragment hits, 11 crystal structures were obtained. This success rate of 20% appears already quite convincing. However, it is difficult to assess whether we missed important fragments, possibly binding in other sub-pockets, to fully map out the rather large binding site of this aspartic protease. Different research groups have suggested the use of alternative prescreening approaches in fragment-based drug discovery in order to limit the risk of losing false negatives or being misled by too many false positives.[11] In practice, such assay strategies are often considered hierarchically. However, a rationale for which assay to use first is not available, and the decision for a particular hierarchy is driven predominantly by the availability of an assay and the effort required to perform the screen. Accordingly, in the present study we evaluated our library with different screening methods in parallel. This allowed us to determine the mutual overlap between the various assay scenarios and to see whether any combination or hierarchical protocol of several assays would have proposed a deviating list of most promising candidates for a potential follow-up crystallographic study.

When screening our 361-entry library using the above-mentioned six different prescreening techniques, we tried to use the most similar assay conditions possible in light of the different requirements for the individual methods (e.g., acidic pH of 4.6; for STD-NMR, the pD was 4.7). However, because the applied prescreening assays operate at somewhat distinct fragment concentrations (0.1 to 2.5 mM, cf. Table 2.1) and show different sensitivity in detecting weak binding, deviating lists of promising hits could be obtained for the crystallographic follow-up analysis. In this context, it is also important to bear in mind that the lower threshold or detection limit to define binding of a promising fragment hit is usually defined based on empirical concepts which are sometimes quite arbitrarily adjusted. In the present study, the assigned hit-definition thresholds are based on practical experience gained in the different contributing groups with the assay systems and are specified in Table 2.1. The description of the conducted experiments can be found in the Experimental Section.

The detailed assay results for 41 fragments that were identified as hits by at least two prescreening methods are listed in Table 2.1. In case we would compile our hit list based on the criterion that a fragment has to show up in only one of the assays, the list would have covered in total 239 compounds corresponding to 66% of the library. Hits are highlighted in green to distinguish them from non-hits.

Unfortunately, not all compounds could be measured in all assays. Special restrictions and applied standards required that in some cases only subsets could be evaluated (Table 2.1).

Therefore, we will consider, apart from the total number of hits, particularly the relative hit rates normalized by the overall number of compounds subjected to the particular assays. Furthermore, three entries were accidentally present twice in the library. This also explains why our library contains 361 entries and not 364 fragments as described in our previous report.[35]

Table 2.1. *Characteristics and hit rates of the applied screening methods.*

Method	c(Fragment)(mM)[a]	c(Protein)(μM)	Buffer	DMSO (% v/v)	# of compounds analyzed (% of full library)	Hit identification	# of hits	Hit rate (%)
Biochem. assay	1.0	0.004	0.1M NaAc pH=4.6	1.0	325 (90%)	>40% inhibition	56[d]	17
RDA	1.7	0.02	0.1M NaAc pH=4.6	0.85	358 (99%)	>30% inhibition	50	14
STD-NMR	0.3	9.6	0.05M CD_3CO_2Na pD=4.7	-	206 (57%)[e]	%STD_AV > 1.75 and %CTL <50	22	11
TSA	2.5	2.2	0.05M Na formate pH=4.6	2.7	330 (91%)	$\Delta T_m \geq 1.0°C = 2.5\ \sigma$	25	8
ESI-MS[b]	0.1	20	0.1M NaAc pH=4.6	1.0	342 (95%)	Categories A+, A and (B)	8 (113)	2 (33)
MST[c]	0.5	0.063-0.125	0.1M NaAc pH=4.6	5.0	282 (78%)[e]	$\Delta F_{norm} \geq 6.0$	36	13
	2.5				279 (77%)[e]		75	27

[a] Fragments for which lower concentrations had to be used due to solubility issues are specified in Table S1. [b] In the course of this study the smaller subset, consisting of categories A+ and A (lacking B), is used. [c] All analyses of this study are based on the MST experiments at 0.5 mm fragment concentration. [d] This number deviates by one additional hit from the 55 hits in the original publication[35] because the raw data were re-evaluated for some fragments, and the hit definition cutoff was adapted from 40% to >40% inhibition. [e] Number of investigated fragments comparably low due to very rigorous fragment quality control process (see Experimental Section).

We took this opportunity and analyzed whether the results produced by the prescreening assays for these identical compounds were similar. Fortunately, there are no major discrepancies between the individual results. The only exception may be that for one duplicate the RDA indicated 40 and 22% inhibition, respectively, which would have led to alternative classifications based on the chosen 30% hit-definition cutoff.

All assays result in a crude relative ranking of the compounds' binding strength, for example, expressed in terms of percent effect. In TSA the amount of thermal stability enhancement can be used to quantify binding; however, we recently showed that an increase in melting temperature highly depends on the adopted binding mode.[39] As this study was also performed on endothiapepsin, we assume that the same effects will hold true for the study of our fragments.

Table 2.2 *Screening data for all fragments that were indicated as hits by at least two methods.* [a]

Fragment	***Biochemical Assay***	***RDA***	***NMR***		***ESI-MS***	***TSA***	***MST***		***Solubility*** [l]
ID	**%inhibition**	**%inhibition**	**%STD_AV**	**%CTL**	**category**[f]	**ΔT_m (°C)**	**ΔF_{norm} (at 2.5 mM)**	**ΔF_{norm} (at 0.5 mM)**	**(µM)**
5	89[d]	34	2.6	37	WB	≤0.8	-1.3	3.8	1846
17	97	68	1.9	76	WB	1.9±0.1[g]	-6.4	-18.0	1445
40	73	46	≤0.5[m]	NB[n]	WB	0.9±0.2[g]	-6.2	-3.8	NA
41	49	36	3.0	26	A	1.6±0.3[g]	2.5	-1.6	>2000
51	46	56	≤0.5[m]	NB[n]	B	≤0.8	169.2	62.5	1674
63	66	37	NA	NA	B	0.6±0.1[g]	NA	-0.6	1860
64	99	16	≤0.5[m]	NB[n]	WB	1.9±0.1[g]	-13.8	-18.5	1937
65	50	31	1.1	29	WB	≤0.8	-5.2	-2.3	1771
66	NA[e]	42	NA	NA	B	1.0±0.1[g]	8.4	-3.0	1843
73	42[d]	31	≤0.5[m]	NB[n]	WB	≤0.8	3.2	0.6	>2000
88	47	35	0.9	90	WB	NA[i]	-11.5[j]	-6.4	1674
109	89	26	2.3	32	WB	≤0.8	-1.5[j]	-6.8	1572
132	<20	-16	≤0.5[m]	NB[n]	B	1.0±0.1[g]	6.7	7.5	>2000
133	28	35	2.8	77	A	≤0.8	-7.1	-2.2	1838
137	47	30	NA	NA	B	≤0.8	55.8[j]	-9.6	686
148	51	18	NA	NA	NS	≤0.8	-5.2	1.8	1950
149	100	77	NA	NA	WB	2.8±0.4[g]	NA	NA	1574
159	60	72	4.0	41	WB	1.9±0.1[g]	-21.2	-13.6	1479
161	80	-25	NA	NA	A	1.1±0.1[g]	-13.6	-5.1	1802
166	<20	44	NA	NA	B	≤0.8	-34.6	-37.3	1595
167	24	46	NA	NA	B	≤0.8	5.3[j]	-34.0	6226
175	75	23	NA	NA	WB*	1.0±0.1[g]	NA	NA	1626
177	100[b]	99	4.8	37	WB	3.4±0.3[g,h]	-16.2	-8.8	1460
178	100[d]	81	2.7	96	WB	1.8±0.1[g]	64.9[j]	-19.2	1916
183	60	16	1.8	64	WB	≤0.8	471.2[j]	-6.1	>2000
196	<20	76	1.9	46	WB	≤0.8	-2.5	0.8	1834
216	45	54	1.9	27	WB	1.8±0.1[g]	3.9	1.6	1965
222	52	30	3.1	32	NS	≤0.8	-3.9	1.0	1175
224	63	87	3.9	38	WB	1.1±0.1[g]	-12.0	-9.2	1806
236	100	75	2.4	153	WB	1.5±0.2[g]	-2.4	1.2	>2000
255	84	36	NA	NA	B	1.9±0.1[g]	-21.3	-19.4	1673

Table 2.2 (cont.) *Screening data for all fragments that were indicated as hits by at least two methods.* [a]

261	47	48	NA	NA	WB	0.7±0.1[g]	-3.7	-4.9	1878
284	87	58	NA	NA	WB	2.0±0.1[g]	-9.0	-3.6	1789
290	84	88	6.7	22	A+	1.2±0.1[g]	-0.7	-1.4	>2000
291	50	-17	2.7	47	B	≤0.8	-11.5	-7.4	>2000
301	75	-120	3.0	33	B	≤0.8	3.5	-1.5	1905
306	93	89	6.7	24	A+	1.7±0.2[g]	-4.0	-2.0	>2000
308	53[d]	35	2.2	42	WB	≤0.8	5.8	57.9[j]	1483
333	76	2	≤0.5[m]	NB[n]	B	<-0.8	-15.3	-7.4	>2000
335	60	25	≤0.5[m]	NB[n]	WB	1.8±0.0[g]	2.1	-0.5	1985
337	38	47	NA	NA	WB	1.3±0.3[g]	NA	NA	>2000
362	43	22	5.5	83	WB	≤0.8	-24.1	-14.7	1348

[a] In addition to all fragments indicated as hit by at least two techniques, the table contains all 11 fragments that were identified as crystallographic hits in a previous study highlighted in the first column in yellow[35] (fragment 148, thus, adds to the table). Hits were defined on the basis of the cutoffs shown in Table 2.1 and are highlighted in green. Fragments that showed a positive result in only one of the two NMR screens or that belong to category B of the ESI-MS assay are indicated in pale green. Fragments with a solubility between1.5 and 2.0 mM are also highlighted in pale green, whereas fragments with solubility ≥ 2.0 mM are shown in green. Finally, fragments for which no data could be recorded or analyzed properly are highlighted in gray. [b] NA: not available due to the absence of significant fragment absorption at wavelengths between 250 and 500 nm. [c] The various ESI-MS categories are defined as follows: A+ :most interesting fragments with high binding extent without significant nonspecific adduction, A: interesting fragments with a relatively high ratio of specific binding to nonspecific binding, B: potentially interesting fragments with a medium ratio of specific binding to nonspecific binding, WB: fragments with low binding extent but without detection of extensive nonspecific adducts, NS: fragments with relatively high binding extent accompanied by high nonspecific adduction(nonspecific binders), WB: fragments with low binding extent (weak binders). [d] Fragment displayed self-fluorescence, but could still be measured. [e] Determined in duplicate by a second and third TSA screen using a temperature increment of 0.2°C min^{-1} (performed for all fragments indicated as hit in the primary screen). All remaining fragments were only investigated in the initial screen (0.5°C min^{-1}), where they were not defined as a hit ($\Delta T_m \leq 2$). [f] Signal-to-noise ratio in STD spectrum too low for integration. [g] NB: non-binder; an exact value could not be calculated in case footnote f applies. [h] Not analyzed because of self-fluorescence. [i] Data could not be analyzed accurately due to the appearance of the melting curve. [j] These data points were excluded from the analysis because the investigated fragments changed the initial fluorescence by > 18 %with respect to the mean initial fluorescence of positive and negative controls on the same capillary chip. [k] Measurement at 0.5 mM fragment concentration. [l] Measurement at 1.25 mM fragment concentration.*

On first glance, it appears more difficult in the case of the thermal shift, ESI-MS and MST assays to distinguish active site binders from “unspecific” surface binders (although in the case of ESIMS and MST[40] specific binding can be assessed through competition assays, provided a high-affinity reference compound is available), while the biochemical assay requires direct binding to the active site to be functional (apart from the unlikely event of an allosteric inhibition in case of the present aspartyl protease). Moreover, in the case of the STD-NMR and RDA, known active site displacement ligands are used to confirm binding to the catalytic site. Here, an appropriate selection of the displacement ligand is crucial. In RDA, the large heptapeptidic ligand pepstatinA was used as reference ligand, whereas in STD-NMR primary screening hits were validated by competitive displacement using ritonavir, a smaller active-site-specific inhibitor with a Ki value

of 15 nm for endothiapepsin.[41] As this inhibitor leaves some sub-pockets in the substrate binding cleft of endothiapepsin unoccupied, the ligand might even provide additional interaction sites for putative fragments and can recruit further fragment hits not found in the primary NMR screen. These facts have to be considered as they may produce non-corresponding results for structural reasons.

In Table 2.1 the total numbers of hits and hit rates resulting from the different assay techniques are summarized. Overall, the biochemical and Proteros RDA assays reveal a similar number of hits. When considering only the pure binding event in STD-NMR (without confirmation via active site displacement, and thus taking the risk of the possibility of unspecific surface binding) the number of hits increases from 22 to 71 (based on the 206 studied compounds). In comparison, the secondary screen using ritonavir as displacement ligand revealed 30 hits. Because this number exceeds the 22 confirmed hits identified in the primary and the ritonavir screen by eight, it could well be that the displacement ligand facilitated the discovery of additional hits. Applying the more restrictive hit-definition conditions (only categories A and A+), the ESI-MS analysis discovers a total of eight strongly binding hits. However, a rather high hit rate of 113 out of 342 investigated fragments is found when choosing the less stringent hit-definition limit (including category B; for details see legend of Table 2.1 and Experimental Section). Using the TSA, 25 hits were identified, corresponding to a 7.6% hit rate. The MST assay was run at two concentrations. While at the lower fragment concentration (0.5 mM) 36 hits were detected, this number increased to 75 fragments at 2.5 mM.

Even though the overall hit rates of the individual screening methods might still be of similar magnitudes (Table 2.1), the mutual overlap between the methods is by no means satisfactory (cf. Venn diagrams in Fig. 2.1 and 2.2). Not a single fragment is characterized as hit by all prescreening techniques (Fig. 2.1). Even when comparing only the four methods with the largest mutual overlaps (biochemical assay, RDA, STD-NMR and TSA; see Table 2.3 as well), the number of only seven common hits is rather disappointing (Fig. 2.2). Furthermore, a comparison of matching hit list pairs reveals that certain techniques such as the biochemical assay and RDA show fairly similar hits, whereas the MST hit set is more distinct from all other screening results (Table 2.3). Interestingly, a total of 119 of our fragments were indicated as hits in at least one assay when only considering ESI-MS categories A+ and A, the 22 competitive NMR binders and the MST results at 0.5 mM fragment concentration (for details see legend of Table 2.1 and Experimental Section). Slightly weakening these criteria by including the MST results at 2.5 mM

fragment concentration leads to 153 hits; adding also the results for both NMR screens individually results in 182 hits. Finally, including the ESI-MS category B as well leads to 239 hits. These large overall hit numbers of 119, 153, 182, and 239 fragments correspond to 33, 42, 50, and 66% of the total library, respectively, again highlighting that the overlap of the different hit lists is rather low. In particular, only 41, 21, 10, 6, or none of the 119 prescreening hits are indicated as binders by at least two, three, four, five, or all six assay techniques, respectively.

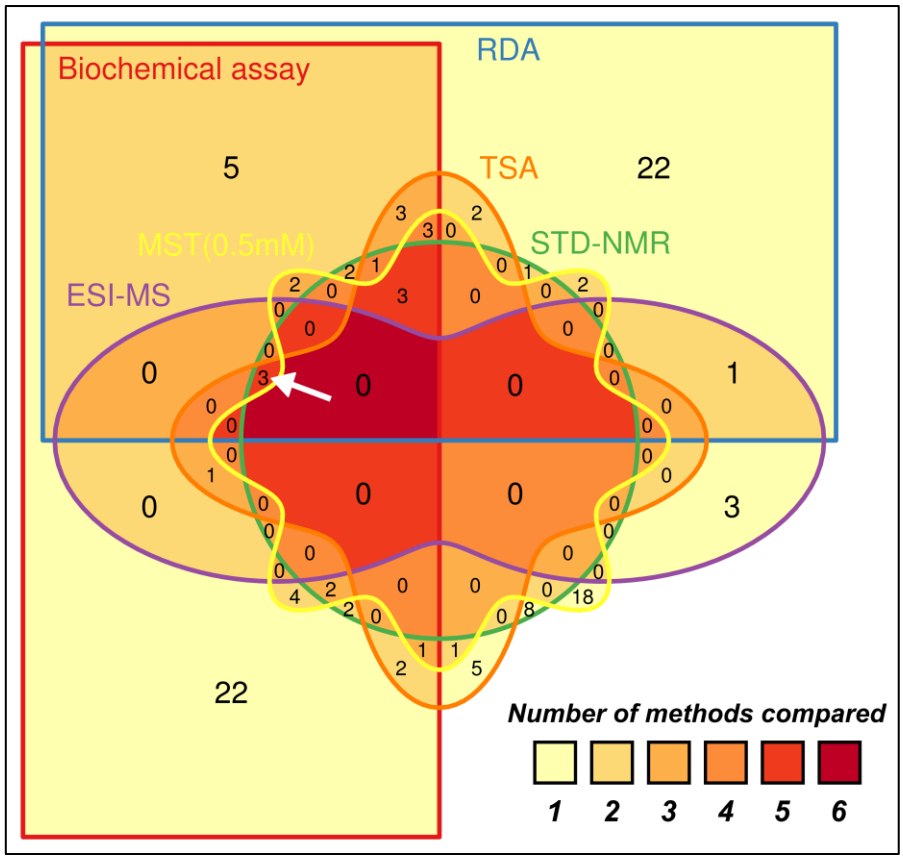


Fig. 2.1. *Venn diagram for all hit sets of the applied prescreening techniques. As indicated by the legend at lower right, the number of overlapping hit sets in a certain region of the diagram is indicated by a gradually increasing reddening of the respective field. All hits revealed by the biochemical assay, RDA, STD-NMR (competitive binders), ESI-MS (categories A+ and A), TSA and MST (at 0.5 mM fragment concentration) are enclosed by red, blue, green, purple, orange, and yellow lines, respectively. For instance, the sum of all numbers within the green circle gives the overall hit count for STD-NMR (24x0+2x1+3x2+2x3+1x8=22). To get the total number of fragments indicated as hits by STD-NMR and RDA (independent of which other method might also have found these fragments as hit), one has to sum over the numbers included in the upper half of the green circle corresponding to the overlap between the green circle and the blue box (11x0+2x1+1x2+2x3=10). The number 3 highlighted by the white arrow represents the number of common hits between five different methods (as indicated by the shade of light red), namely the biochemical assay, RDA, STD-NMR, ESI-MS, and TSA (but not MST, as the object surrounded by the yellow line does not include this number). This graphic was produced using the statistical framework R and Inkscape.*

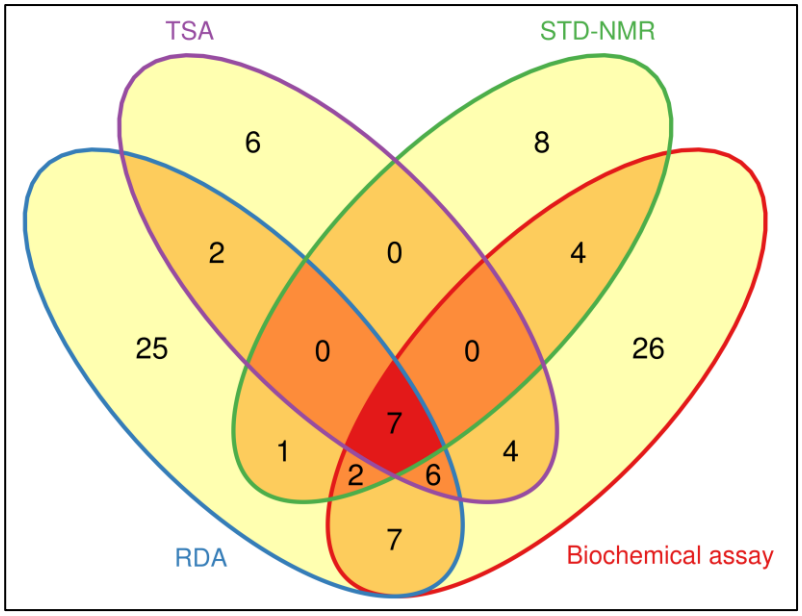


***Fig. 2.2**. Venn diagram for those prescreening hit sets with the largest mutual overlaps. The number of overlapping hit sets in a certain region of the diagram is highlighted by a gradually increasing color intensity. All hits of the biochemical assay, RDA, STD-NMR (competitive binders) and TSA are enclosed by red, blue, green, and purple lines, respectively. This Venn plot was generated by R.*

***Table 2.3**. Number and percentage of overlapping hits between pairs of the applied screening methods [a]*

Method	Biochem. assay	RDA	STD-NMR[b]	TSA	ESI-MS[c]	MST[d]
Biochem. assay	x	22 (39%)	13 (23%)	17 (30%)	4 (7%)	15 (27%)
RDA	22 (44%)	x	10 (20%)	15 (30%)	4 (8%)	10 (20%)
STD-NMR[b]	13 (59%)	10 (45%)	x	7 (32%)	3 (14%)	5 (23%)
TSA	17 (68%)	15 (60%)	7 (28%)	x	4 (16%)	8 (32%)
ESI-MS[c]	4 (50%)	4 (50%)	3 (38%)	4 (50%)	x	0 (0%)
MST[d]	15 (42%)	10 (28%)	5 (14%)	8 (22%)	0 (0%)	x

[a] The percentage given in parentheses reflects the fraction of all hits identified by a certain method that is confirmed as hit by the technique depicted in the respective column; for example, 44% of all RDA hits are also found in the biochemical assay. [b] Only the 22 competitive binders were taken into account. [c] The small subset consisting of categories A+ and A (lacking B) was used for this analysis. [d] MST results from the measurements at 0.5 mM fragment concentration were analyzed.

Despite the unsatisfactory consensus of the applied screening methods, the obtained results appear rather convincing with respect to the 11 determined crystal structures as, overall, the RDA, NMR, MS, TSA, and MST analyses would have suggested 10 of the actually detected crystallographic hits. In particular, the RDA, NMR, MS (excluding/including category B), TSA, and MST (at 0.5/2.5 mM fragment concentration) assays would have been able to identify 73, 100, 27/55, 55, and 27/33%, respectively, of the crystallographic hits found owing to the evidence from our biochemical assay. For NMR, however, the result must be interpreted based on the fact that with this method only seven of the 11 crystallographic binders could be investigated. When decreasing the size of the library to the 55 fragments investigated crystallographically and taking the resulting X-ray hits obtained after the biochemical assay as a benchmark, a crystallographic follow-up screen with the 21- fragment hit list indicated by the RDA would have resulted in eight structures. Analogously, the NMR, ESI-MS (excluding category B), thermal shift, and MST (at 0.5 mM fragment concentration) assays would have revealed 7, 3, 6, or 3 structures out of 13, 4, 17, and 15 prescreening hits, respectively. However, conclusions about whether some screening methods might be superior to others regarding the prediction of crystallographic binders have to be taken very carefully, as our crystallographic hits were only found after applying the biochemical prescreening filter, which might, for instance, penalize the MST results because they have a comparably low overlap with the biochemical assay hits (Table 2.3). Possibly applying a different prescreening protocol or combining several of them could have detected another set of fragment complex structures.

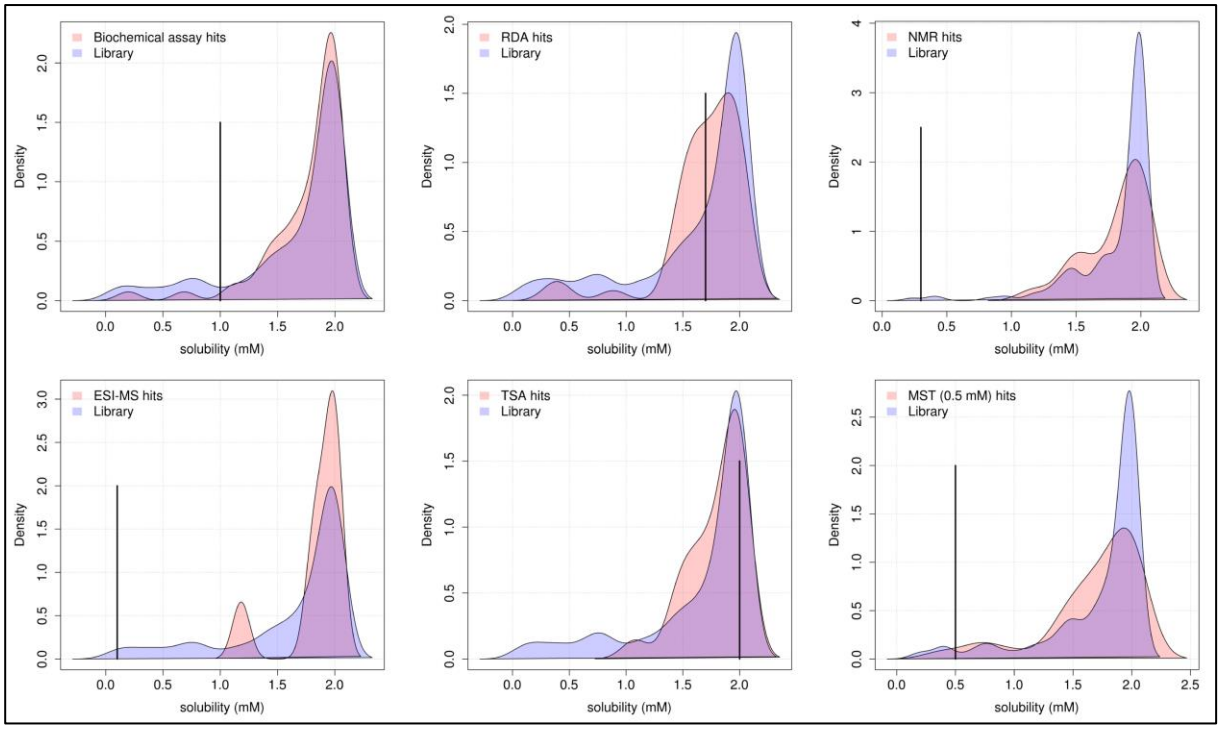


***Fig. 2.3**. Distribution of measured solubility with respect to the applied screening methods. As reference, the fragment screening concentration of each method is highlighted by a vertical line (see Table 2.1 and the Experimental Section for details regarding experimental conditions). Each plot compares the solubility kernel density estimations of all library fragments that were investigated by a specific screening technique (blue) with those indicated as a hit by the same method (pink). All values indicating at least 2 mM solubility were summarized in the 2 mM bin. Therefore, the reference line for the TSA, which was conducted at a 2.5 mM fragment concentration, was also drawn at 2.0 mM. The depicted plots were generated using the statistical framework R.*

Moreover, it is very informative to compare the measured fragment solubility of all library entries investigated by a certain screening method against the suggested list of hits (for solubility values see Table 2.2). The corresponding solubility distributions displayed in Fig. 2.3 clearly highlight that most hits display a solubility above ~1 mM, whereas fragments with a solubility <1 mM are less likely to be identified as hit. One exception of this observation, however, was the MST assay, where also for less soluble compounds binding was equally probable. Nevertheless, our results suggest that sufficient solubility is usually an important property of prescreening hits and that it would therefore be wise to omit insufficiently or less soluble compounds during the design of new fragment libraries. For instance, in our case this has been achieved for the library subset investigated by NMR (cf. Fig. 2.3, first row, and third column). The entire library mostly contains fragments with a solubility >1 mM (85 %). The fragment concentration used for most of the screens remained below this limit (Table 2.1 and Fig. 2.3). Thus, the majority of the fragments were fully soluble under the applied assay conditions. Only in the reporter displacement and thermal shift assays screening concentrations >1 mM were applied. Nevertheless, these methods

also detected fragment hits with solubility below the applied screening concentration. To avoid artifacts possibly resulting from compound aggregation, it is advisable to screen at the lowest possible concentrations that are still high enough with respect to the sensitivity of a given method.

We also compared the methods that more likely detect specific active site binding (biochemical assay, RDA) with those also observing binding events remote from the substrate binding pocket (ESI-MS, TSA, MST). In this comparison, the NMR assay played a crucial role, as this method is, in principle, capable of discriminating specific active site binding from unspecific binding, with the additional effect that the displacement ligand ritonavir may recruit further fragment binders. To perform the latter analysis, we divided the NMR hit list into ritonavir- competitive, ritonavir-uncompetitive, and ritonavir-induced binders (Fig. 2.4). As anticipated, the functional assay hits (biochemical assay, RDA) match better as active site binders with the ritonavir-competitive binders than with the putative remote binders, whereas the globally sensitive methods (ESI-MS, TSA, MST) display a larger overlap with the ritonavir-uncompetitive binders. Interestingly, the latter methods only retrieved one of the putative ritonavir-induced binders, underlining that the presence of the displacement ligand might indeed be necessary for binding to take place in these cases. However, further experiments will be required to confirm that specific fragments can actually only bind in the presence of ritonavir. An important question remains: why do we detect such a limited overlap among the different fragment screening approaches? This observation is shared by other researchers in the field as communicated at meetings; however, little results have been thoroughly documented on this fact in the literature, or the assays were performed under conditions that are difficult to compare.[42] The biophysical effects exploited by the various methods are quite distinct, which raises the question how much overlap actually can be expected from diverse screening techniques. Furthermore, any assay requires the choice of a distinct hit-definition threshold necessarily influencing the screening results as well as specific conditions to be fulfilled. We tried to keep parameters such as pH, salt conditions, and ionic strength similar or at least in a comparable range, but as the influence of these parameters on binding is multifactorial along with the differences in the methods' sensitivity, the overall effects on the screening results are extremely difficult to estimate. Moreover, it has to be regarded that fragment screening operates at rather high concentrations where micelle formation or aggregation of organic compounds, fragment–substrate interactions, interference with fluorescence detection (absorption, quenching, auto-fluorescence) or partial protein denaturation

can occur and might affect the effective fragment and/or protein concentration differently for distinct screening techniques.

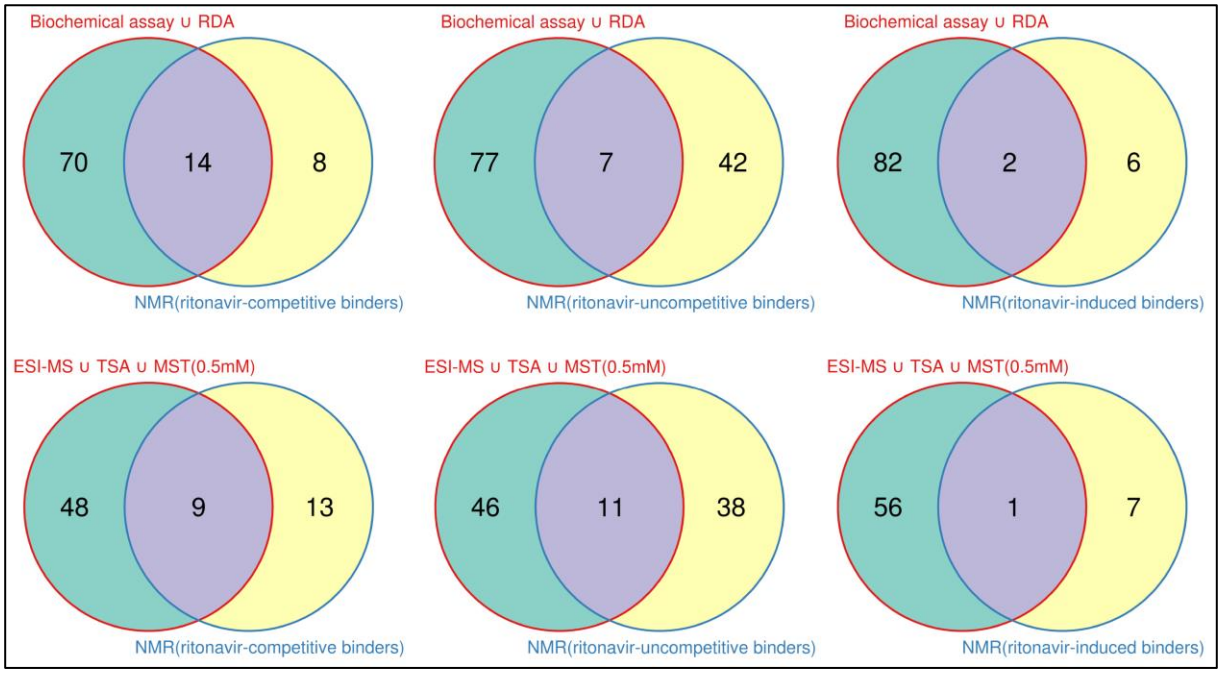


Fig. 2.4*. Comparison of screening methods with respect to their ability to detect active site and unspecific binders. The first row contains three Venn diagrams contrasting the merged biochemical assay/RDA hit set (green) with those NMR hits (yellow) that were obtained in the primary and secondary (left, competitive binders with respect to ritonavir), in the primary but not in the secondary (center, uncompetitive binders with respect to ritonavir), and in the secondary but not in the primary screen (right, fragments binding presumably due to the presence of ritonavir). In the second row, the same NMR categories are compared with the set union of the ESI-MS (categories A+ and A), TSA and MST (at 0.5 mM fragment concentration) hits. All Venn diagrams were produced using R.*

2.5 Conclusions

The last 15 years of fragment-based lead discovery have demonstrated that the approach is only really powerful if the actual binding mode of a fragment in the protein binding pocket is known, particularly when working with proteins harboring very large pockets. Therefore, a crystallographic analysis of fragment binding will be essential in the end. The present study shows, admittedly based on the experience with one protease as a reference, that all prescreening methods produce different results with minor overlap of putative hits. Usually, crystallography is not taken as a primary screening technique as it is considered to be too expensive, elaborate, time-consuming and cumbersome. But at the end, the question arises whether the yield detected through one of the rapid and well established prescreening methods provides enough hits for a follow-up crystallographic study or whether possibly some important, perhaps better suited, or

even patent-free fragments are missed. Apart from taking this putative risk, from a drug designer's point of view every structure of a crystallographically confirmed fragment hit provides important insight in terms of an experimental active site mapping of the studied target protein. Even if such a fragment is finally not taken as the candidate for hit-to-lead optimization, the knowledge of its binding mode might help to create new ideas at a later stage of a drug development project. This can be of utmost importance to make a drug discovery program finally successful.

2.6 Experimental Section

2.6.1 Protein material

Endothiapepsin samples were obtained from Suparen (provided by DSM Food Specialties); the buffer was exchanged with 0.1M sodium acetate (pH 4.6) using a Vivaspin 20 with a 10 kDa molecular weight cutoff. Protein concentration was determined by absorbance at λ = 280 nm, assuming an extinction coefficient of 1.15 for 1 mg·mL^{-1} solutions.[43]

2.6.2 Fragment library

The fragment library was assembled and purchased as previously described.[35] The originally reported compound collection contained three redundant entries, which were independently evaluated in most assays providing the opportunity to check for inconsistencies and to assess the reproducibility of the assay results. The solubility of all fragments was determined in buffered solution as described elsewhere.[35]

2.6.3 Biochemical assay

All 361 entries of the library were tested against endothiapepsin in a functional assay using the peptide substrate Abz-Thr-Ile-Nle-(*p*-nitro-Phe)-Gln-Arg-NH_2.[35] Upon substrate cleavage, fluorescence increased during the measurement. The assay was performed in a 0.1M sodium acetate buffer at pH 4.6 containing 0.01% (v/v) Tween 20. Substrate and the investigated screening compound were pre-incubated and the protein was added directly before measurement. The final reaction mixture contained 4 nM endothiapepsin, 1.8 µM substrate, 1 mM test compound, and 1% (v/v) DMSO. Under these conditions, 18 test compounds were insufficiently

soluble and 18 additional compounds had to be studied at a concentration of 500 µM due to limited solubility. Finally, 17 fragments had to be excluded from the assay due to self-fluorescence. Each compound was measured twice, and the percent inhibition given in Table 2.2 represents the average of both measurements; 56 fragments led to enzyme inhibition of >40%.

2.6.4 Thermal Shift assay

345 sufficiently soluble fragments were subjected to a thermal stability assay following the protocol we described previously.[39] Samples were heated from 20 to 80°C at a rate of 0.5°Cmin^{-1}. Of the 345 melting curves, 15 could not be analyzed accurately due to their appearance. Potential hits (ΔT_m > 0.8°C) were re-measured twice at a rate of 0.2°C min^{-1}. The 0.1M sodium acetate assay buffer at pH 4.6 contained 0.01% (v/v) Tween 20 and a mixture of 2.2 µM endothiapepsin, the probe fragment at 2.5 mM (1.25 mM for the less soluble fragments), and 0.2% (v/v) of the dye SYPRO Orange to record unfolding (final DMSO concentration: 2.7% (v/v)). Shifts of the melting temperature with respect to native endothiapepsin were determined based on the analysis of the first derivative of the recorded fluorescence/temperature protocol.

Thermal shifts of $\Delta T_m \geq 1.0$°C, which corresponds to the 2.5-fold standard deviation of 15 T_m determinations of the *apo*-enzyme in the presence of DMSO, were assumed significant, leading to 25 hits.

2.6.5 Native ESI-MS assay

For ESI-MS experiments, the buffer of endothiapepsin samples was exchanged against 50 mM ammonium formate (pH 4.6). The final protein–ligand mixtures contained 100 µM fragment and 20 µM protein in 50 mM ammonium formate (pH 4.6), 5% (v/v) ethanol and 1% (v/v) DMSO. These samples were then analyzed using an automated chip-based nanoelectrospray robot (Triversa Nanomate, Advion Biosciences) connected to an electrospray time-of-flight mass spectrometer (LCT, Waters) operating with fine-tuned instrument parameters promoting the detection of intact noncovalent complexes.[44] Under these conditions, 342 of the 361 fragments could be investigated. Compounds were ranked based on relative complex abundances, applying penalties when nonspecific binding was suspected from the detection of

statistically distributed multi-adduction of the compound to the protein. This resulted in a sixfold classification scheme. The two best-ranked compound categories, A+ and A, correspond to

compounds with relatively high binding extent with no or low nonspecific adduction, respectively. Intermediate categories B (compounds with medium binding extent and some nonspecific binding) and WB* (weak binders without nonspecific adduction) encompass compounds for which no clear conclusion could be drawn regarding classification as true or false positive hits. Finally, less interesting compounds were classified in category NS (compounds showing extensive nonspecific adduction) and WB (weak binders showing some nonspecific adduction). Two entries were top-ranked (category A+, 0.6% of database referring to 342 entries), six fragments fell into the second category (category A, 1.8%) and 105 compounds (30.7%) were classified as potential binders with medium, slightly nonspecific binding (category B). For our comparative analysis we decided to consider the eight top-ranked compounds of the first two categories, and the 105 hits from category B were additionally regarded.

2.6.6 Reporter displacement assay

The reporter displacement assay (RDA), developed at Proteros, has been described in detail previously.[19,20] In brief, the reporter displacement assay is based on a reporter ligand that is distinctively designed to bind to the site of interest of the chosen target enzyme or receptor. The proximity between reporter ligand and protein results in the emission of an optical signal. Fragments and compounds that bind to the same binding site displace the reporter ligand to cause a signal loss. Quantity and kinetics of signal loss can be monitored. The reporter displacement assay is a homogeneous method performed in 384-well plates allowing the measurement of 384 kinetic traces in parallel. In the present case, pepstatin A, a potent and crystallographically confirmed inhibitor of endothiapepsin,[45] was used as reporter ligand. The 361 fragments were tested for their binding to the protease. In the case of competitive binding to the pepstatin A binding site, the test fragments displaced the reporter ligand, resulting in loss of the fluorescence signal. Due to the nature of the optical signal, interference with the fluorescence properties of the fragments can occur. The measurements were performed at a fragment concentration of 1.7 mM and a protein concentration of 20 nM (final DMSO concentration: 0.85% (v/v)) in a sodium acetate buffer at pH 4.6. Overall, 358 of the 361 fragments could be investigated, and the results are reported as percent inhibition of the protein.

For 50 fragments, binding led to an inhibition of >30% under the applied assay conditions. To quantify the inhibitory potential of the various fragments, the amount of signal loss was quantified at different fragment concentrations once the system had reached equilibrium. Following this protocol, IC_{50} values were determined for the 25 strongest binders, which are summarized in Table 2.4.

***Table 2.4**. IC_{50} values for the 25 top-ranked fragment hits from the RDA.*

Fragment	IC_{50}[mM]	Fragment	IC_{50}[mM]	Fragment	IC_{50}[mM]
5	> 1.30	148	> 1.30	261	0.59
6	0.23	149	0.15	284	0.26
17	0.10	152	> 1.30	285	> 1.30
33	1.66	177	0.003	290	0.04
40	2.09	178	0.23	291	> 1.30
41	> 1.30	196	0.81	306	0.04
63	2.28	216	0.09	327	0.10
75	> 1.30	224	0.02		
78	> 1.30	255	1.16		

2.6.7 Microscale thermophoresis (MST) assay

A subset of 295 fragments, mainly selected based on their DMSO solubility (>50 mM), was analyzed in an MST assay. Endothiapepsin was labeled with the Monolith NT.115 Protein Labeling Kit RED-NHS from NanoTemper Technologies (NanoTemper Technologies, Munich, Germany). For labeling, the endothiapepsin stock was diluted in PBS (pH 7.0) to reach a final concentration of 20 µM in a total volume of 100 µL. A pH of 7 is required in this particular case to prevent endothiapepsin inactivation which occurs at pH values >7. Endothiapepsin solution was then mixed with 100 µL PBS (pH 7.0) containing 200 µM NT647-NHS fluorophore and incubated at room temperature for 15 min. Free fluorophore was separated from labeled protein by size-exclusion chromatography using gravity-flow size-exclusion columns provided in the Protein Labeling Kit RED-NHS (NanoTemper Technologies).

For assay development, potential adsorption of NT647–endothiapepsin to the capillary walls was investigated in pre-tests with 60 nM NT647–endothiapepsin in assay buffer (100 mM sodium

acetate (pH 4.6), 0.05% (v/v) Tween 20, 5% (v/v) DMSO) in standard treated and premium-coated MST-grade capillaries (NanoTemper Technologies) using a Monolith NT.115 instrument (NanoTemper Technologies). Because the protein did not adsorb to the two capillary types under these conditions, all subsequent experiments were performed in standard treated capillaries. The interaction of NT647–endothiapepsin with the peptidomimetic inhibitor acetylpepstatin was used to identify optimal assay conditions and was used as a positive control throughout the screening. Quantitative binding experiments using a 16-fold 1:1 serial dilution of acetylpepstatin showed that the optimal binding signal was obtained at an MST power of 80% with quantification of thermophoresis after 25 s, yielding a K_d value of ~150 nM and a binding amplitude of 15.9 counts, while the baseline noise of NT647–endothiapepsin was 3 counts under these conditions. The same interaction was measured by MST after 1 h incubation at room temperature, yielding similar results, showing that the interaction was stable during this time span.

For fragment screening, single-point measurements were performed in assay buffer. Two different fragment concentrations (0.5 and 2.5 mM) were used to investigate the influence of this parameter on the resulting hit list for this rather new technique in the FBLD field. While samples containing DMSO without fragments were used as negative controls, acetylpepstatin was present at 120 µM in the positive controls. The labeled protein was added to the reaction mixture to yield a final concentration of 62.5–125 nM. MST measurements were conducted using the Monolith NT.Automated instrument (NanoTemper Technologies) at 80% laser power, 40–50% LED power, and 25°C. F_{norm} values were derived by dividing the fluorescence intensity after an MST laser on-time between 25 and 26 s and the initial fluorescence prior to MST laser activation. ΔF_{norm} values were then obtained by subtracting F_{norm} values of the negative controls from F_{norm} values in the presence of the respective fragments on the same capillary chip. Fragments that altered the initial fluorescence by >18% relative to the average initial fluorescence of negative and positive controls on the same capillary chip were excluded from the analysis (0.5 mM: 13 fragments excluded, 2.5 mM: 15). MST data were then exported from the NT Affinity Analysis and plotted in Spotfire software (Tibco, Palo Alto, CA, USA) for evaluation of normalized fluorescence values (F_{norm}) versus initial fluorescence. Runs at 0.5 and 2.5 mM fragment concentration were evaluated individually. A fragment was defined as a hit if the $|\Delta F_{norm}|$ diverged by >2-fold from the total averaged negative control noise (3 F_{norm} counts) to the respective negative control on the same capillary chip. This analysis revealed 36 hits at a fragment concentration of 0.5 mM and 75 hits at 2.5 mM.

2.6.8 STD-NMR assay

For the NMR experiments (Bruker Avance II, 600 MHz with cryogenic probe head) only fragments that meet the following three criteria were considered: 1) HPLC–MS purity ≥80%, 2) aqueous solubility in the applied buffer of at least 500 µM, and 3) no aggregation (as measured by a self STD intensity <2%) at 500 µM. This resulted in the selection of a subset of 206 fragments for this comparative study. To save data acquisition time, mixtures of two compounds were tested. The first set of NMR measurements (binding experiments) was performed in D_2O buffer solution containing 50 mM CD_3CO_2D with 100 mM NaCl and 4 mM NaN3, adjusted to pD 4.7 (a pH of 4.3 was determined) at 25°C. Sample mixtures contained 310 µM of each fragment compound and 9.6 µM endothiapepsin. The total assay volume of 163 µL was filled in 2.5 mm NMR tubes. To distinguish specific interactions from nonspecific binding outside the active site, competitive displacement experiments were performed with all fragment mixtures in the presence of 16 µM ritonavir as a potent active site inhibitor.[41] Individual fragment signals were identified by comparing the screening spectra with the two corresponding reference spectra, which were recorded separately in the absence of endothiapepsin or ritonavir. For each binding fragment, averaged STD-NMR intensity values were determined by comparing the difference with the off-resonance spectrum (Table 2.1, STD_AV>1.75 %). This was done for the binding and the corresponding competitive displacement experiment, respectively. Percentage control values (CTL [%]) were calculated by dividing the STD intensity values from the binding experiment and the competitive displacement experiment (% CTL = 100*(*I*(competitive displacement)/*I*(binding))). Fragments with an averaged percent CTL of <50% were classified as competitive binders. Overall, 22 competitive binders were identified by STD-NMR referring to the reduced sample set of 206 compounds.

2.7 References

[1] Carr, R. A.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-based lead discovery: Leads by design. *Drug Discovery Today* **2005**, *10*, 987–992.

[2] Schulz, M. N.; Hubbard, R. E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.* **2009**, *9*, 615–621.

[3] Chessari, G.; Woodhead, A. J. From fragment to clinical candidate--a historical perspective. *Drug Discovery Today* **2009**, *14*, 668–675.

[4] Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **2012**, *51*, 4990–5003.

[5] Erlanson, D. A., Jahnke, W., Eds.; Methods and principles in medicinal chemistry 34; John Wiley distributor: Weinheim, Chichester, 2006.

[6] Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.* **2004**, *47*, 3463–3482.

[7] Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008**, *51*, 3661–3680.

[8] Mattos, C.; Ringe, D. Locating and characterizing binding sites on proteins. *Nat. Biotechnol.* **1996**, *14*, 595–599.

[9] Hilpert, K.; Ackermann, J.; Banner, D. W.; Gast, A.; Gubernator, K.; Hadvary, P.; Labler, L.; Mueller, K.; Schmid, G. Design and Synthesis of Potent and Highly Selective Thrombin Inhibitors. *J. Med. Chem.* **1994**, *37*, 3889–3901.

[10] Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.

[11] Mashalidis, E. H.; Sledz, P.; Lang, S.; Abell, C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat. Protoc.* **2013**, *8*, 2309–2324.

[12] Proll, F.; Fechner, P.; Proll, G. Direct optical detection in fragment-based screening. *Anal. Bioanal. Chem.* **2009**, *393*, 1557–1562.

[13] Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.

[14] Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR* **2001**, *21*, 349–359.

[15] Dalvit, C.; Fagerness, P. E.; Hadden, D. T. A.; Sarver, R. W.; Stockman, B. J. Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.* **2003**, *125*, 7696–7703.

[16] Dalvit, C. NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discovery Today* **2009**, *14*, 1051–1057.

[17] Śledź, P.; Lang, S.; Stubbs, C. J.; Abell, C. High-Throughput Interrogation of Ligand Binding Mode Using a Fluorescence-Based Assay. *Angew. Chem. Int. Ed.* **2012**, *124*, 7800–7803.

[18] Chen, L.; Cressina, E.; Leeper, F. J.; Smith, A. G.; Abell, C. A fragment-based approach to identifying ligands for riboswitches. *ACS Chem. Biol.* **2010**, *5*, 355–358.

[19] Neumann, L.; Ritscher, A.; Muller, G.; Hafenbradl, D. Fragment-based lead generation: identification of seed fragments by a highly efficient fragment screening technology. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 501–511.

[20] Neumann, L.; Konig, K. von; Ullmann, D. HTS reporter displacement assay for fragment screening and fragment evolution toward leads with optimized binding kinetics, binding selectivity, and thermodynamic signature. *Methods Enzymol.* **2011**, *493*, 299–320.

[21] Lo, M.-C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **2004**, *332*, 153–159.

[22] Kranz, J. K.; Schalk-Hihi, C. Protein thermal shifts to identify low molecular weight fragments. *Methods Enzymol.* **2011**, *493*, 277–298.

[23] Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R. High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *J. Biomol. Screening* **2001**, *6*, 429–440.

[24] Hofstadler, S. A.; Sannes-Lowery, K. A. Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. *Nat. Rev. Drug Discovery* **2006**, *5*, 585–595.

[25] Vivat Hannah, V.; Atmanene, C.; Zeyer, D.; van Dorsselaer, A.; Sanglier-Cianferani, S. Native MS: an 'ESI' way to support structure- and fragment-based drug discovery. *Future Med. Chem.* **2010**, *2*, 35–50.

[26] Ganem, B.; Li, Y. T.; Henion, J. D. Detection of noncovalent receptor-ligand complexes by mass spectrometry. *J. Am. Chem. Soc.* **1991**, *113*, 6294–6296.

[27] Navratilova, I.; Hopkins, A. L. Fragment screening by surface plasmon resonance. *ACS Med. Chem. Lett.* **2010**, *1*, 44–48.

[28] Frostell-Karlsson, Å.; Remaeus, A.; Roos, H.; Andersson, K.; Borg, P.; Hämäläinen, M.; Karlsson, R. Biosensor Analysis of the Interaction between Immobilized Human Serum Albumin and Drug Compounds for Prediction of Human Serum Albumin Binding Levels. *J. Med. Chem.* **2000**, *43*, 1986–1992.

[29] Neumann, T.; Junker, H.-D.; Schmidt, K.; Sekul, R. SPR-based fragment screening: advantages and applications. *Curr. Top. Med. Chem.* **2007**, *7*, 1630–1642.

[30] Silvestre, H. L.; Blundell, T. L.; Abell, C.; Ciulli, A. Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 12984–12989.

[31] Ruhmann, E.; Betz, M.; Fricke, M.; Heine, A.; Schafer, M.; Klebe, G. Thermodynamic signatures of fragment binding: Validation of direct versus displacement ITC titrations. *Biochim. Biophys. Acta* **2015**, *1850*, 647–656.

[32] Jerabek-Willemsen, M.; Wienken, C. J.; Braun, D.; Baaske, P.; Duhr, S. Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **2011**, *9*, 342–353.

[33] Blundell, T. L.; Jhoti, H.; Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discovery* **2002**, *1*, 45–54.

[34] Jhoti, H.; Cleasby, A.; Verdonk, M.; Williams, G. Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Curr. Opin. Chem. Biol.* **2007**, *11*, 485–493.

[35] Koster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011**, *54*, 7784–7796.

[36] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discovery Today* **2003**, *8*, 876–877.

[37] Schuffenhauer, A.; Ruedisser, S.; Marzinzik, A. L.; Jahnke, W.; Blommers, M.; Selzer, P.; Jacoby, E. Library design for fragment based screening. *Curr. Top. Med. Chem.* **2005**, *5*, 751–762.

[38] Siegal, G.; Ab, E.; Schultz, J. Integration of fragment screening and library design. *Drug Discovery Today* **2007**, *12*, 1032–1039.

[39] Kuhnert, M.; Koster, H.; Bartholomaus, R.; Park, A. Y.; Shahim, A.; Heine, A.; Steuber, H.; Klebe, G.; Diederich, W. E. Tracing binding modes in hit-to-lead optimization: chameleon-like poses of aspartic protease inhibitors. *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 2849–2853.

[40] Seidel, Susanne A. I.; Dijkman, P. M.; Lea, W. A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J. S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; Katritch, I.; Melo, F. A.; Ladbury, J. E.; Schreiber, G.; Watts, A.; Braun, D.; Duhr, S. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **2013**, *59*, 301–315.

[41] Behnen, J.; Koster, H.; Neudert, G.; Craan, T.; Heine, A.; Klebe, G. Experimental and computational active site mapping as a starting point to fragment-based lead discovery. *ChemMedChem* **2012**, *7*, 248–261.

[42] Wielens, J.; Headey, S. J.; Rhodes, D. I.; Mulder, R. J.; Dolezal, O.; Deadman, J. J.; Newman, J.; Chalmers, D. K.; Parker, M. W.; Peat, T. S.; Scanlon, M. J. Parallel screening of low molecular weight fragment libraries: do differences in methodology affect hit identification? *Journal of Biomolecular Screening* **2013**, *18*, 147–159.

[43] Larson, M. K.; Whitaker, J. R. Endothia parasitica Protease. Parameters Affecting Stability of the Rennin-like Enzyme. *J. Dairy Sci.* **1970**, *53*, 262–269.

[44] Sanglier, S.; Atmanene, C.; Chevreux, G.; van Dorsselaer, A. Nondenaturing mass spectrometry to study noncovalent protein/protein and protein/ligand complexes: technical aspects and application to the determination of binding stoichiometries. *Methods Mol. Biol.* **2008**, *484*, 217–243.

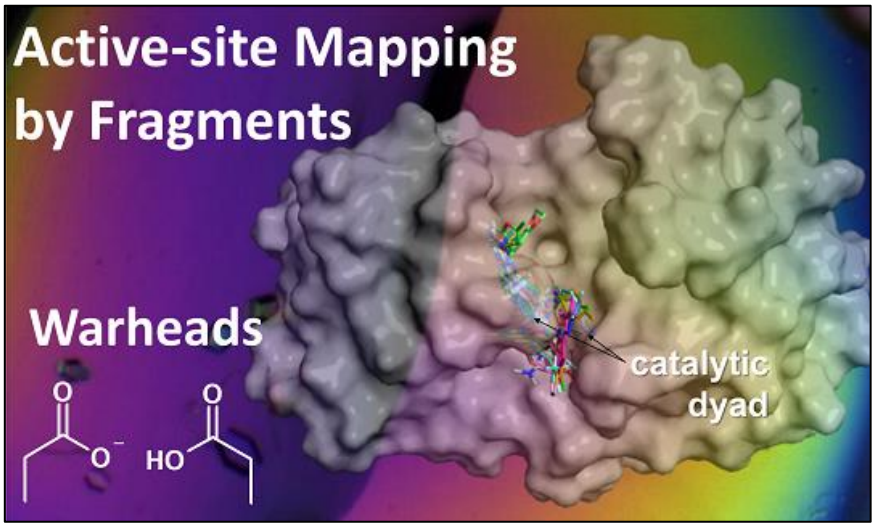
[45] Pearl, L.; Blundell, T. The active site of aspartic proteinases. *FEBS Letters* **1984**, *174*, 96–101.

3. Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads To Address a Catalytic Dyad

3.1 Introductory Remarks

The following chapter has been published in *J. Med. Chem.* in 2016. The chapter is a part of a major project including the fragment soaking, and data collection of over 450 datasets, discussed in the following chapter and in chapter 4. Protein purification, crystallization and all soaking experiments were performed by the author of this thesis with additional help from Johan Winquist and Ah Young Park. For the data collection and structure refinement, the work of six additional colleagues, Johannes Schiebel, Frederik R. Ehrmann, Stefan G. Krimmer, Kan Fu, Xiaojie Wang, and Martin Stieler was consulted. The data collection was done at Helmholz-Zentrum Berlin für Materialien und Energie, BESSY II with the help of Franziska U. Huschmann, Manfred S. Weiss, and Uwe Müller. All data collection and refinement statistics can be found in the Appendix.

The author of this thesis carried out the refinement of all datasets collected for fragments **F001**–**F070** and performed the data analysis and interpretation for all fragment hits along with drafting the paper.



*Reprinted with permission from *J. Med. Chem.* **2016**, *59*, 9743–9759. Copyright 2016 American Chemical Society.

3.2 Abstract

Crystallography is frequently used as follow-up method to validate hits identified by biophysical screening cascades. The capacity of crystallography to directly screen fragment libraries is often underestimated, due to its supposed low-throughput and need for high-quality crystals. We applied crystallographic fragment screening to map the protein-binding site of the aspartic protease endothiapepsin by individual soaking experiments. Here, we report on 41 fragments binding to the catalytic dyad and adjacent specificity pockets. The analysis identifies already known warheads but also reveals hydrazide, pyrazole, or carboxylic acid fragments as novel functional groups binding to the dyad. A remarkable swapping of the S1 and S1′ pocket between structurally related fragments is explained by either steric demand, required displacement of a well-bound water molecule, or changes of trigonal-planar to tetrahedral geometry of an oxygen functional group in a side chain. Some warheads simultaneously occupying both S1 and S1′ are promising starting points for fragment-growing strategies.

3.3 Introduction

Over the past 10 years fragment-based lead discovery (FBLD) has become popular and is increasingly used as first step in the lead discovery process.[1,2] The fundamental idea of this approach is to start with small (MW < 250 Da) but efficiently binding seed fragments, which are subsequently grown to lead molecules optimized for a given target protein based on structural information obtained from X-ray crystallography.[3] Usually, at the initial stage of FBLD, biophysical methods ("prescreening methods" in the following) such as native mass spectrometry (MS), thermal-shift assay (TSA), saturation transfer difference (STD) NMR, surface plasmon resonance (SPR) or microscale thermophoresis (MST) are used to screen a fragment library.[4−7] These methods are applied to retrieve the most promising fragment binders from a library.[8,9] Although they are expected to be sensitive enough to identify weak binding compounds, none of them can unambiguously determine the adopted binding mode of discovered hits. However, structural knowledge is an essential prerequisite for subsequent rational hit-to-lead optimization. Furthermore, it is rarely considered that prescreening is often equally laborious, expensive, time consuming and in fact frequently leads to unsatisfactorily overlapping hit lists.[10] A serious limitation for the applicability of many of the suggested techniques is the compound solubility in aqueous buffer, a problem often occurring when high concentrations of weak binding compounds

are required for the analysis. An alternative to overcome this bottleneck is the application of X-ray crystallography as screening method right at the beginning of an FBLD campaign. This strategy allows much higher concentrations than are usually applicable in any other technique, often accomplished by adding organic solvents such as DMSO to the crystallization buffer.[11−13] The capability of crystallographic fragment screening largely remains underexploited, although it is one of the most sensitive techniques considered for rational fragment screening.[14,15] We therefore decided to crystallographically screen our entire in-house fragment library in order to thoroughly explore the binding site of the aspartic protease endothiapepsin (EP) by elucidating different binding geometries.

Our previously presented library of 361 fragments had been designed using a set of predefined criteria, slightly extending the rule-of-three concept for physicochemical properties.[16,17] The chosen protein EP[18] has been used in the past as a surrogate for aspartic proteases of the A1 family, e.g., drug targets such as renin (hypertension), plasmepsins (malaria), β-secretase (Alzheimer's disease), or secreted aspartic proteinases (fungal infections).[19−21] EP shares folding architecture and considerable sequence identity with its related targets and represents a robust crystallization system producing well-diffracting crystals. However, in a recent study we demonstrated that EP is still a challenging target, as minor structural modifications of a parent lead molecule resulted in significant changes in binding modes.[22] Nonetheless, we believe that EP is a relevant model protein for method development in basic drug research, as it provides high resolution structural data often not achievable with other aspartic proteases.

All 361 fragments were soaked individually into premanufactured crystals of the *apo* protein. This is a difference in strategy compared to our previous study,[17] where we preselected only promising fragment candidates according to a biochemical assay and applied cocktails of the detected hit compounds. To reach equilibrium for weaker and potentially slow-binding fragments, a rather long soaking time of 48 h was used, which was well tolerated by the crystals. In total, we identified 71 hits bound to the target protein. As one of these seemed to be oligomerized, 70 hits were kept for the subsequent analysis. Of these, 41 hits bound to the catalytic dyad, corresponding to a hit rate of 11% (Fig. 3.1). An additional 29 fragment hits were found to bind remotely from the catalytic dyad. These are discussed in a companion paper with respect to the binding geometry and drug design concepts.[23] Overall, 86 distinct binding poses could be characterized as some fragments occurred multiple times in the binding pocket.

According to the proposed enzyme mechanism, the catalytic aspartates (Asp35 and Asp219) have varying pK_a values due to the influence of their local environment.[24,25] At the beginning of the catalytic cycle, Asp35 is believed to be deprotonated and charged, while Asp219 is protonated.[26–28] Thus, all fragments directly binding to the catalytic dyad require hydrogen-bond acceptor and/or donor functionalities to interact with the aspartates.

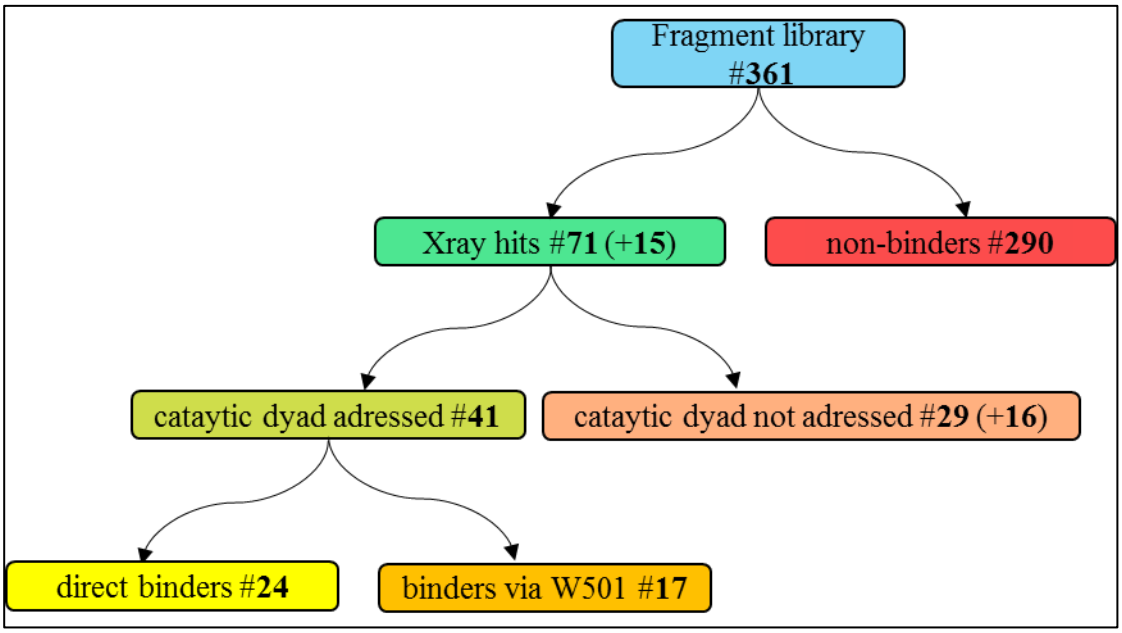


Fig. 3.1*. Results for the crystallographic fragment screening using our 361 fragment library. The number of detected hits is given with respect to their binding mode and spatial location on the target protein. Several fragment molecules were found with at least one additional copy, resulting in 86 distinct binding poses (41 + 29 + 16). Of these, only four hits were found outside the substrate-recognition site. The 41 hits addressing the catalytic dyad were divided into direct and water-mediated binders.*

Subsequently we describe a detailed comparative analysis of all 41 individual fragment hits. We believe that the obtained knowledge of their exact binding mode to the target is essential to understand the caused effect, allowing us to make informed design suggestions for potency increase and ultimate hit-to-lead optimization.

3.4 Results

The analyzed crystal structures of the 41 fragments have a mean resolution of 1.3 Å and overall completeness of 98.5%, highly important criteria for unambiguous fragment assignment in the electron density. The data collection yielded high quality characteristics with a mean R_{sym} of 5.7%, $\langle I/\sigma(I)\rangle$ of 16.6, and Wilson B factor of 10.4 Å^2 (c.f. Appendix, Table 8.1).

Overall, we were able to produce high-quality data perfectly well suited for subsequent hit-to-lead optimization and other methodological investigations.

3.4.1 Binding Sites and Overall Physicochemical Properties of Fragment Hits.

Fragment hits were divided into two main categories: (1) catalytic dyad binders, mainly located next to the S1 and S1′ pockets and (2) remote binders, which populate the extended binding site of EP, distal to the catalytic dyad (cf. chapter 4).[23] The catalytic dyad binders scatter across the active site, occupying the various recognition pockets S1 to S3′ (Fig. 3.2).

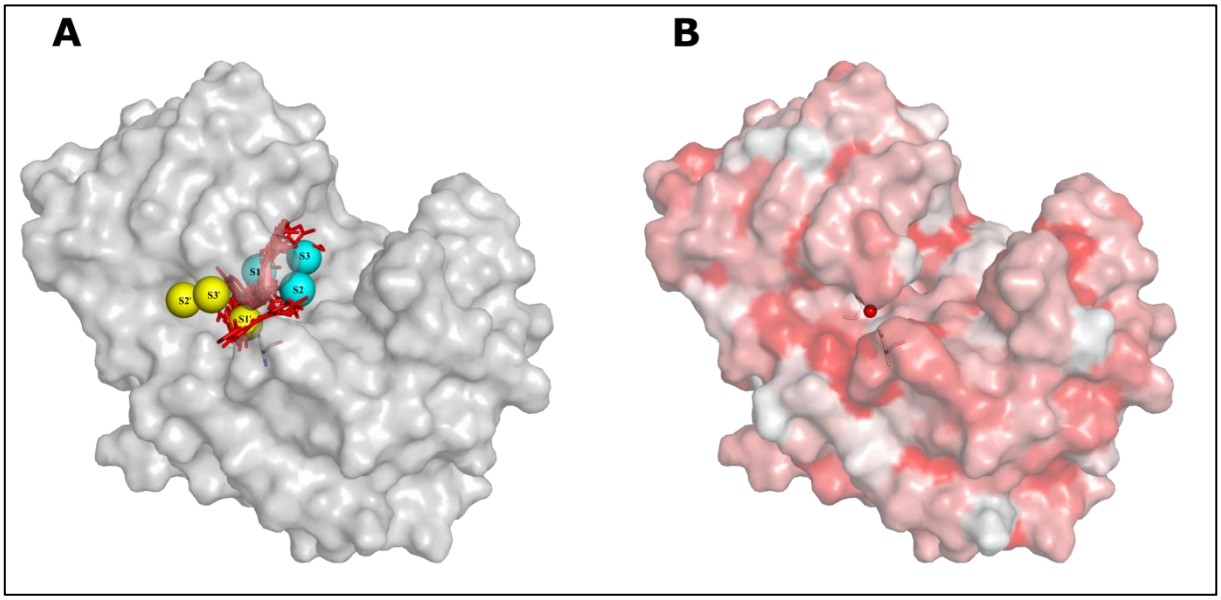


***Fig. 3.2**. (A) Superimposition of apo EP (PDB code 4Y5L) with all 41 direct fragment binders, depicted as red sticks, showing their spatial distribution across the EP binding site. The recognition pockets S3′ through S3 are indicated by yellow and cyan spheres (cf. Experimental Section). (B) Protein surface color-coded by hydrophobicity scale. Red intensity of surface patches indicates increasing hydrophobicity according to Eisenberg and Schwarz.[29]*

All relevant pockets will be described in the following section. As part of the flap region, the polar Ser78, Tyr79, and Gly80 residues are located close to pocket S3′. The S2 pocket is also surrounded by polar residues, including Thr222, Thr223, and Tyr226. The S3 pocket, flanked by Ile10, Ala16, and Ile122, is partly covered by the flap region, and the S1 pocket, surrounded by Tyr79, Phe116, and Leu125, is completely buried under the flap. The flap region furnishes the S3 and S1 pockets with amphiphilic character. The S2′ pocket is rather hydrophobic, as it is surrounded by Ile77, Leu133, and Phe194. Overall, S1′ is the most hydrophobic pocket, being largely surrounded by Ile300, Ile302, and Ile304.

With respect to the 41 fragments that bind to the catalytic dyad of EP we next analyzed whether certain physicochemical properties allow distinguishing X-ray hits from nonbinders. Due to the protonation state of its catalytic site, EP has a strong preference to bind positively charged fragments at pH 4.6 used in all crystallographic experiments (Fig. 3.3). Considering only those 101 fragments from our library that are likely positively charged at pH 4.6, our screen would have resulted in 47 fragment-bound structures, corresponding to a hit rate of 47%. This preference for positively charged fragments is a result of the acidic properties of EP (pI = 4.3 in the folded state according to the PDB2PQR (www.poissonboltzmann.org/docs/structures-ready/)),[30] which renders its active site negatively charged even at the applied pH of 4.6. Since the catalytic Asp35 is assumed to be negatively charged as supported by neutron diffraction studies,[31] it may not come as a surprise that 13 of all 68 aliphatic amines in our library (19%) turned out to be catalytic dyad binders in our crystallographic screen. The hit rate is even higher for catalytic dyad binders containing aliphatic primary amines (9 of 20 = 45%), whereas it is slightly lower for aliphatic secondary amines (3 of 23 = 13%) but equal to zero for 33 tertiary amines. Likewise, significant enrichments were found for amidines and guanidines (9 of 23 = 39%) as well as for aromatic amines (9 of 56 = 16%). The obvious preference of EP for the binding of primary over secondary and tertiary amines is likely caused by the limited space available next to the deeply buried catalytic dyad, making it difficult to accommodate bulkier amines.

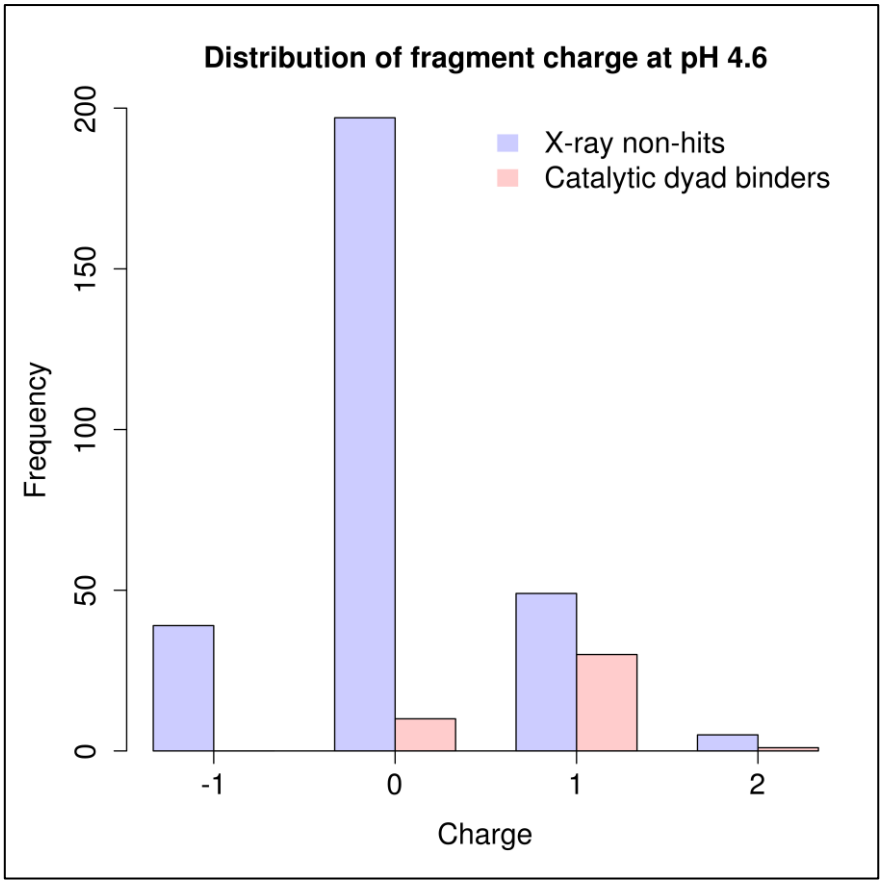


Fig. 3.3*. Distribution of charges for the 41 X-ray hits binding to the catalytic dyad (pink) compared to the 290 nonbinders (blue) estimated at pH 4.6 (cf. Experimental Section). Since the p-value is below 0.05 (pt test = 4 × 10−12, pMann−Whitney U-test = 8 × 10−13), the observed difference between binders and nonbinders is highly significant. This plot was generated using R.*[32]

In addition to charge, we investigated other physicochemical properties important for fragment binding and compared the 41 catalytic dyad hits to all nonbinders (Table 3.1). Notably, there is no significant difference between the X-ray hits and nonbinders with regard to their average number of rotatable bonds (2.6 vs 2.7) or total polar surface area (TPSA: 51 vs 52 Å2). However, the X-ray hits have on average a 30 Da lower molecular weight compared to the nonbinders, underlining that fragments with smaller molecular weight or fewer non-hydrogen atoms have a higher propensity to bind to the active site, agreeing well with the key concepts of FBLD.[33] The 41 catalytic dyad binders also possess on average a higher number of hydrogen-bond donors (2.0) than the nonbinders (1.2). The most probable reason for this is the preference of EP to bind nitrogen-containing, basic, and thus hydrogen-bond donating molecules.

***Table 3.1**. Comparison of physicochemical properties between the 290 non-binders[1] and 41 catalytic dyad binders.*

	Non-binders		**Catalytic dyad binders**		***p*-value**[(a)]	
	Range	*Average*	*Range*	*Average*	*Independent samples t-test*	*Mann-Whitney U-test*
No. of non-H atoms	8 - 20	15.0 ± 3.2	8 - 20	13.5 ± 2.9	0.003	0.004
MW (g/mol)	128.2 - 353.3	222.0 ± 43.1	122.2 - 288.4	192.6 ± 41.0	$8 \cdot 10^{-5}$	$5 \cdot 10^{-5}$
clogP(o/w)	-1.25 - 5.39	1.48 ± 0.89	-0.60 - 3.86	1.20 ± 0.85	0.050	0.015
No. of Lipinski donors	0 - 4	1.2 ± 0.7	0 - 4	2.0 ± 1.0	$3 \cdot 10^{-5}$	$2 \cdot 10^{-6}$
No. of Lipinski acceptors	1 - 7	3.7 ± 1.0	2 - 6	3.4 ± 1.0	0.025	0.025
No. of rotatable bonds[(b)]	0 - 10	2.7 ± 1.6	0 - 6	2.6 ± 1.4	0.615	0.795
TPSA (Å^2)	15.3 - 92.7	52.0 ± 13.3	29.0 - 99.4	51.0 ± 12.4	0.643	0.306

[(a)]p-values indicate whether the difference between catalytic dyad binders and non-binders is significant ($p < 0.05$)
[(b)]Definition according to Oprea considering also ring bonds.[34]

3.4.2 Fragments Binding Directly to the Catalytic Dyad.

Of the 41 X-ray hits, 24 fragments bind directly to the dyad and displace the catalytic water molecule W501. They predominantly occupy the S1 and S1′ pockets. We decided to group the fragment hits by their chemical composition in order to highlight the multiplicity of interaction patterns generated with the carboxylate groups of both aspartates. To allow for a consistent comparison of the conservation of hydration sites, we assigned a uniform numbering scheme to all hydration sites and observed water molecules. The assigned numbers increase with the distance of a considered solvation site with respect to the catalytic W501 as a common reference point (cf. Experimental Section). All water molecules are referred to by this number preceded by "W", while individual fragments are referred to by their library identification number preceded by "F" (fragment identifier).

Hydrazides. Six hydrazides (**F004**, **F075**, **F081**, **F103**, **F109**, and **F330**) bind to the catalytic aspartates via their two hydrazide nitrogen atoms (Fig. 3.4). In **F075**, the N′-substituted hydrazide unit is located between two ring moieties. In five fragments, the hydrazide carbonyl is directly attached to a phenyl ring, which is replaced by a benzyl group in **F103**. The benzene rings are substituted either in meta position by a methyl (**F004**) or a dimethylamino group (**F081**) or in para position by a bromine (**F075**), a fluorine (**F103**), or a diethylamino group (**F109**).

Only **F330** has a chlorine substituent attached in the ortho position. The in-chain hydrazide **F075** exhibits an additional pyrrole-type ring in the N′ position, which is buried in the S1 pocket. While **F004**, **F075**, **F081**, **F103**, and **F330** place their aromatic rings into the S1′ pocket, **F109** unexpectedly adopts a significantly different orientation positioning its benzene ring into the S1 pocket and its tertiary amine close to Phe116 and Ile122. Since **F109** differs from **F081** only by a sterically more demanding para diethylamino instead of a meta dimethylamino substituent, steric crowding might be an explanation for this swapped orientation (cf. Discussion section). While the α-nitrogen of F109 forms one 2.8 Å H-bond to Asp35, the terminal nitrogen atom serves as a donor addressing both Asp35 and Asp219. In contrast, the other five hydrazides form an H-bond of 2.4−2.8 Å between Asp219 and the α-nitrogen as well as bifurcated H-bonds of 2.7−3.1 Å between Asp35 and their terminal nitrogen. While F004 and F081 lack additional interactions with the enzyme, **F075** and **F103** accept a 3.0 Å H-bond from Gly80-NH of the flap region via their carbonyl oxygen.

Primary Amines. Three fragments (**F005**, **F189**, and **F274**) bind to the catalytic dyad via their primary amino groups (Fig. 3.5). Although **F005** has an amidine functionality, its binding mode is similar to that of primary amines. **F005** places the benzo group of its 2H-isoindole moiety into the S1 pocket. As suggested by a predicted pK_a value beyond 9.0 (cf. Experimental Section), the endocyclic nitrogen of **F005** is most likely not protonated and thus accepts an H-bond from Thr222OH located at the far end of the S2 pocket. The exocyclic amino group donates one H-bond to each of the aspartates of the catalytic dyad (3.0 and 3.1 Å). In addition, **F189** occupies the S1 pocket with its pyridine moiety forming a 2.7 Å H-bond between its meta pyridine nitrogen and Asp81 of the flap region. **F189** adopts two alternative conformations (occ. = 42% and 58%) as suggested by the residual difference electron density after placement of the first conformer.

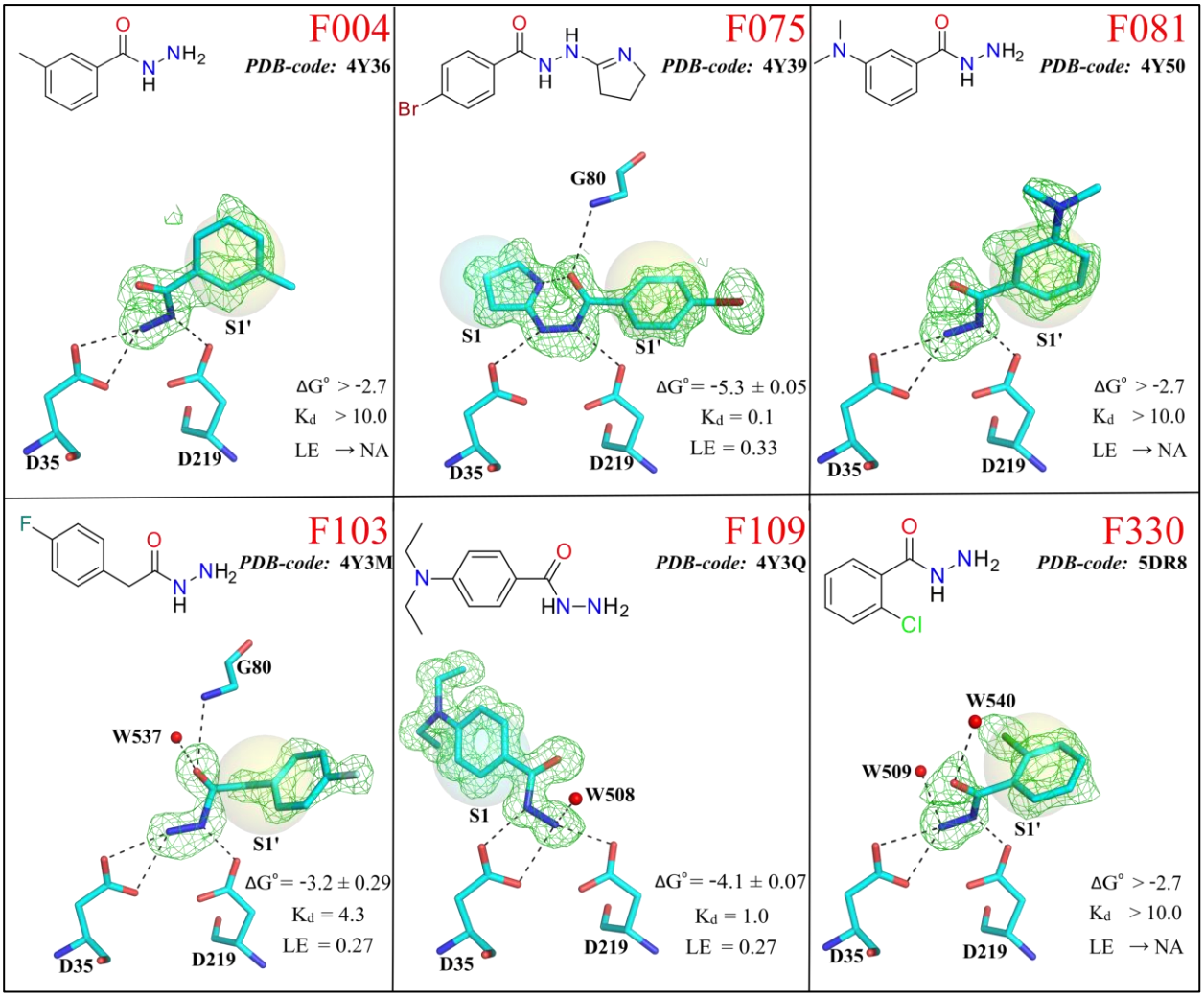


Fig. 3.4*. Binding modes and affinity data of the hydrazides. The fragments are labeled by their library entry number. In all cases, the $mF_o - DF_c$ electron density is shown at 2.5σ, and hydrogen bonds are shown as dotted lines. The preferentially occupied pockets are indicated by translucent yellow and blue spheres. While heteroatoms are indicated by standard color-coding, the carbon atoms have been colored based on the assigned compound class. For all experiments the chiral fragments were used as racemates, but the stereoisomer identified in the crystal structure is depicted. Chemical formulas and PDB codes are shown together with the corresponding fragment crystal structure. The binding affinity obtained by ITC is given as free energy of binding (ΔG^0) in kcal/mol, and K_d values are in mM. The ligand efficiency (LE) was calculated as $-\Delta G^0$/number of heavy atoms and given in kcal/(mol·atom). For very weak binding fragments (K_d > 10 mM), the obtained experimental error was ≥0.5 kcal/mol, thus no exact affinity and LE values are listed.*

In contrast, **F274** positions its para fluorophenyl moiety into the S1′ pocket. The fluorine atom is oriented toward Ile300, Ile302, and Ile304, forming a hydrophobic environment. Notably, even though the shape and van der Waals radius of fluorine (1.47 Å) closely mimic a hydrogen atom,[35] a certain amount of energy must be spent to desolvate the polar fluorine atom in order to accommodate the ligand in the target pocket. In the case where the desolvation energy becomes too large, the binding process will be unfavorable.

The protonated primary amino group of **F274** forms three charge-assisted H-bonds with the catalytic dyad with appropriate lengths of 2.8−3.2 Å. Furthermore, the hydroxyl group at the S-stereocenter accepts an H-bond (2.9 Å) from the Gly80NH of the flap region. In addition, two DMSO molecules occupy the S2 and S2′ pockets, one of which is neighboring the fragment at a distance of 2.6 Å.

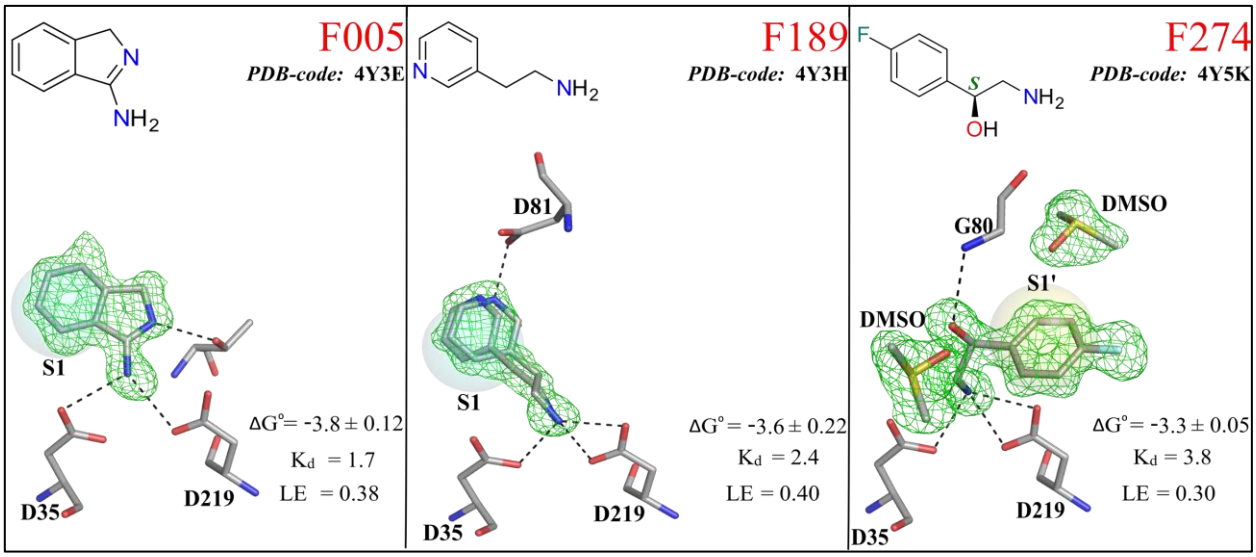


***Fig. 3.5**. Binding modes and affinity data of the primary amines. Representation as in Fig. **3.4**.*

Secondary Amine. Fragment **F236** carries three substituents at its central phenyl moiety: a methylamino-methylene, an ortho chloro, and a meta nitro group (Fig. 3.6). The phenyl moiety occupies the S1 pocket forming a π−π-stacking interaction with Tyr79. The secondary amino group donates one charge-assisted H-bond to each of the catalytic aspartates, as suggested by the predicted pK_a value of 8.4. Additionally, a presumably weak H-bond (3.3 Å) is formed between the nitro group and W569, which in turn forms an H-bond (2.8 Å) to the carbonyl oxygen of Gly221.

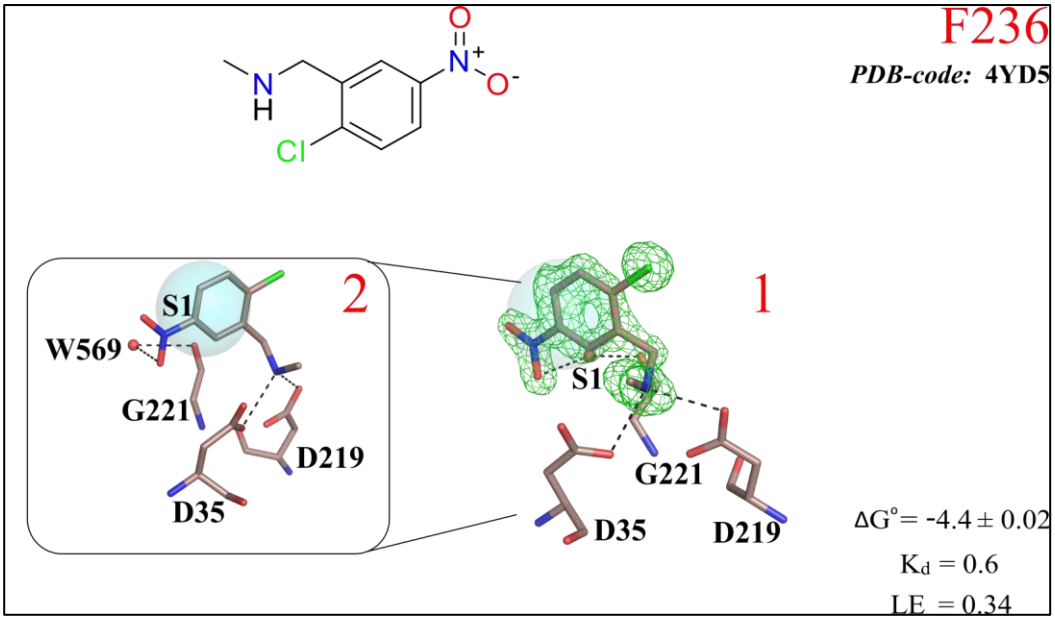


***Fig. 3.6**. Binding mode and affinity data of the secondary amine. Representation as in Fig. **3.4**.*

Piperidine. Three molecules of **F207** bind simultaneously to the EP binding pocket, one of which is only partly localized in the electron density (Fig. 3.7). The fragment consists of a polar piperidin-3-yl moiety connected via an amide group to a second piperidine moiety. The first of the three molecules occupies the S1′ pocket (occ. = 100%) while the other two occupy the S2 (occ. = 86%) and S6 (occ. = 90%) pockets, respectively. The first molecule forms two H-bonds of equal length (2.8 Å), in each case between its protonated piperidin-3-yl and one of the catalytic aspartates. In addition, it establishes an H-bond between its amide oxygen and the Gly80NH of the flap region.

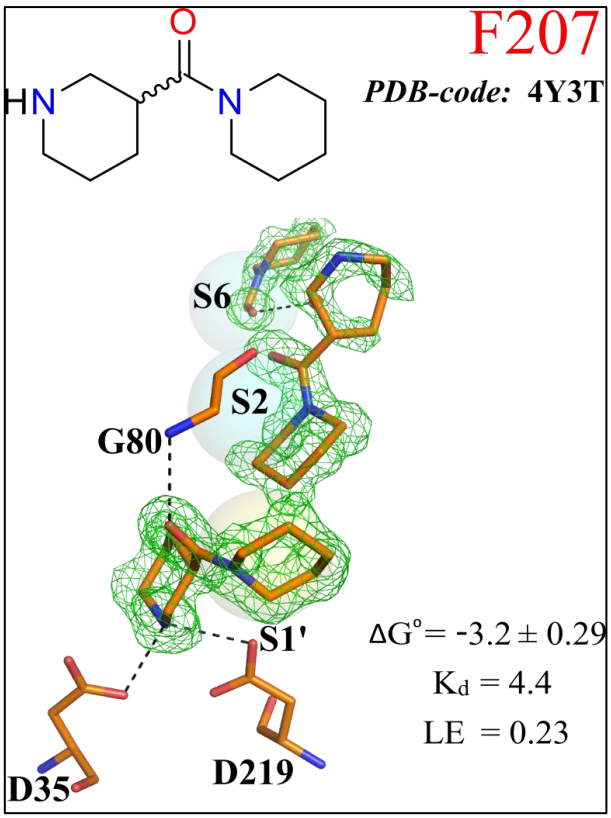


Fig. 3.7*. Binding mode and affinity data of the piperidine. Representation as in Fig.* ***3.4****.*

Guanidine, **Amidine**, **and Isothiourea**. All amidine-type fragments form bifurcated H-bonds to the catalytic dyad. The singly bound **F218** molecule (Fig. 3.8, upper row) contains a terminal guanidino group attached to a quinazolin-2-yl moiety, which is oriented toward the flap region. Notably, seven amino acids from this region could not be modeled into the electron density (Ser78, Tyr79, Gly80, Asp81, Gly82, Ser83, and Ser84), as they would clash with the quinazoline moiety if they adopted their usual conformation observed in all other structures. The most likely charged guanidino group forms an elaborate chelatetype H-bonding network with the catalytic aspartates, in which the nonterminal nitrogen addresses Asp35 indirectly via W512 (2.5 Å).

The three amidines **F125**, **F205**, and **F216** bind directly to the catalytic dyad, most likely in their protonated form (Fig. 3.8, upper and bottom row). While only a single molecule of the para trifluoromethyl benzamidine **F216** is bound in the crystal structure, a second molecule of the amidinopiperidine **F125** and the thiophenamidine **F205** bind simultaneously in the respective structures. The amidino groups of the dyad-contacting molecules form three H-bonds to the aspartates, with **F205** forming an additional H-bond to a nearby glycerol molecule (2.9 Å).

The glycerol is absent in the complexes with **F125** and **F216** where it is replaced by water molecules. The benzene rings of this class of fragments occupy the S1′ pocket locating the para trifluoromethyl group of **F216** in the hydrophobic region formed by Ile300, Ile302, and Ile304. In contrast, the additional fragment molecules of **F125** and **F205** not interacting with the dyad are buried in the S1 pocket, forming a single H-bond to Asp81 of the flap region, with **F205** forming an additional bifurcated H-bond to Ser115.

The two isothiourea fragments **F290** and **F306** show a similar binding pattern to the amidine-type fragments, as they only differ by the inserted sulfur atom next to the amidino group (Fig. 3.8, bottom row). Both **F290** and **F306** are structurally very similar, varying only in the attachment point of the chlorine substituent at the benzyl portion. The para chlorine atom of **F290** is oriented toward the solvent, while the ortho chlorine atom of **F306** forms an edge-on chloro−π interaction with Phe194. Both fragments fill the S1′ pocket with their hydrophobic benzyl moiety facing the binding region composed of Ile300, Ile302, and Ile304. Remarkably, similar to the amidines, the isothiourea fragments form a bifurcated H-bond to Asp35 and an additional H-bond to Asp219.

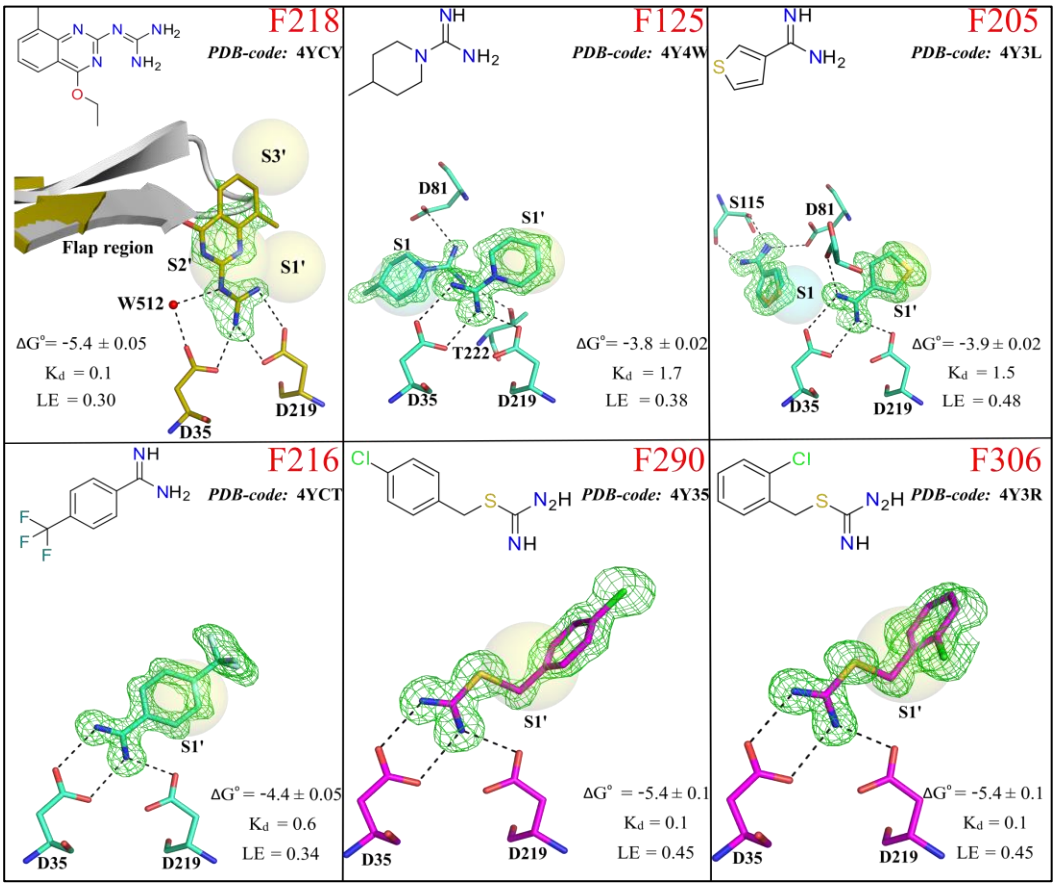


Fig. 3.8*. Binding modes and affinity data of the guanidine, amidine, and isothioureas. Representation as in Fig.* ***3.4****.*

Tetrahydroimidazotriazine, Imidazole, Pyridine, and Pyrazole. Fragment **F054** contains an imidazotriazine heterocycle, which is connected via a single bond to a 3,4-difluorophenyl ring. Two fragment molecules are found in the crystal structure (Fig. 3.9, upper row). The first molecule occupies the S1′ pocket, whereas the second one is accommodated by the S3 and S4 pockets. The five-membered ring of the fused heterocycle does not interact with the catalytic dyad. Instead, the most likely protonated guanidine-type triazine nitrogen donates an H-bond to Asp219 (3.3 Å) and accepts an H-bond from Thr222OH (3.2 Å). The non-guanidine type nitrogen forms most likely as an acceptor a weak H-bond (3.1 Å) to W529. Additionally, water-mediated contacts to Gly80NH (3.2 Å) and Asp81NH (3.1 Å) are formed, both provided from the flap region. A sufficient distance between both fragment molecules of more than 5 Å is given so that a mutual interaction can be excluded.

Fragment **F338** positions its methylated cyclohexyl moiety into the S2 pocket, while the likely protonated imidazole donates an H-bond to Asp35 (2.8 Å) in the S1 pocket (Fig. 3.9, upper row). The second aspartate is too far away to interact with this fragment. Furthermore, the imidazole ring forms a π−π-stacking interaction with Tyr79 (4.1 Å). The amide carbonyl oxygen of the fragment accepts an additional H-bond from Gly80NH, whereas the amide nitrogen interacts with Thr223NH mediated by W575.

A slightly different binding mode is observed for the chiral fragment **F162** (Fig. 3.9, upper row), of which only the *S*-enantiomer was identified in the crystal structure. The pyridine ring is buried in the S1 pocket, benefiting from a π−π-interaction with Tyr79. Under the applied soaking conditions, the pyridine nitrogen (pK_a = 4.8) is likely partially protonated in aqueous solution and might adopt full protonation in its bound state next to the charged Asp35 (2.7 Å). The remaining part of the fragment occupies the S1′ pocket, while the thiophene ring points toward the solvent. In addition, the amide carbonyl oxygen of the fragment accepts an H-bond from Gly80NH (2.8 Å) while the amide nitrogen forms an H-bond to Thr222OH (2.9 Å).

The pyridine ring of **F209**, similar to that of **F162**, interacts with the charged Asp35 at a distance of 3 Å (Fig. 3.9, bottom row). Due to the lack of sufficiently defined difference electron density, the isoxazole moiety of F209 was not modeled in the crystal structure.

Fragment **F255** donates two direct H-bonds to the catalytic dyad, one from the protonated imidazopyridine nitrogen to Asp35 (2.6 Å), the other from the amide nitrogen to Asp219 at a distance of 2.7 Å (Fig. 3.9, bottom row).

The terminal phenyl ring occupies the S1′ pocket and orients toward the solvent, while the imidazopyridine moiety fills the S1 pocket forming a π–π-stacking interaction with Tyr79. Furthermore, the carbonyl oxygen accepts an H-bond from Gly80NH of the flap region.

Two molecules of the pyrazole fragment **F114** bind to EP, one directly to the dyad (Fig. 3.9, bottom row) and a second completely remote on the EP surface. The terminal pyrrolidine ring of the dyad binder occupies the S1 pocket while the substituted pyrazole orients toward the S1′ pocket, directly interacting with the catalytic dyad. Additional interactions are formed between the nitrile nitrogen and Asp81NH as well as between the exocyclic amino group and Thr222OH.

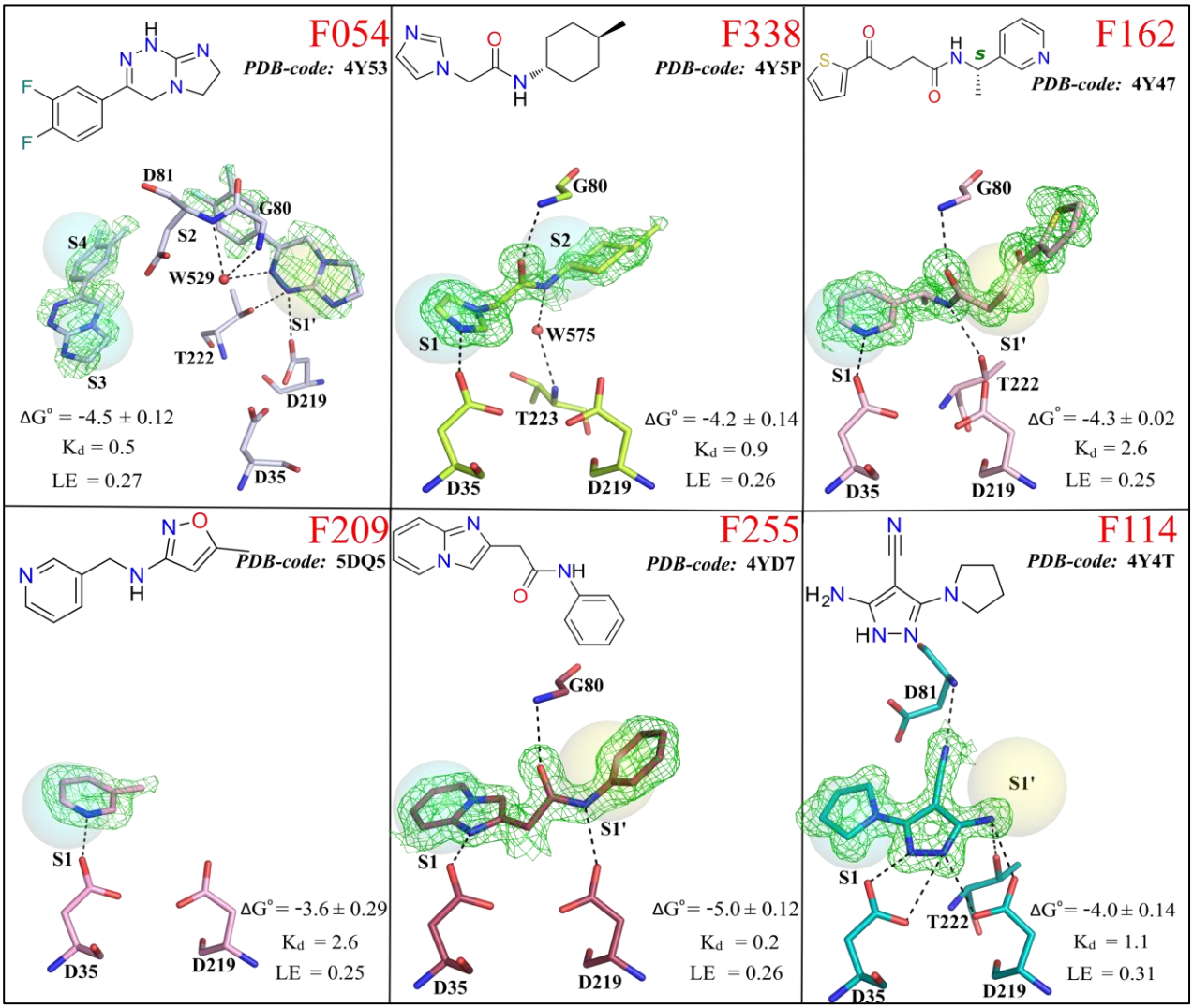


Fig. 3.9*. Binding modes and affinity data of the Tetrahydroimidazotriazine, Imidazole, Pyridine, and Pyrazole. Representation as in Fig.* ***3.4****.*

Alcohol. Not only amino groups but also a hydroxyl group binds to the catalytic dyad directly, as found for the pyridopyrimidine **F231** (Fig. 3.10). The presumably protonated nitrogen is oriented toward Thr222 and W550 forming weak H-bonds (3.4 and 3.3 Å). However, the difference electron density for one ring carbon atom was insufficiently defined, suggesting the binding of an open-chain structure in the crystal structure. This could be due to the formation of a ketone tautomer under the applied acidic soaking conditions. Considering the H-bond distances of the fragment to the catalytic dyad, we assume that the flexibility of the nonaromatic ring led to the ill-defined density around the missing carbon atom.

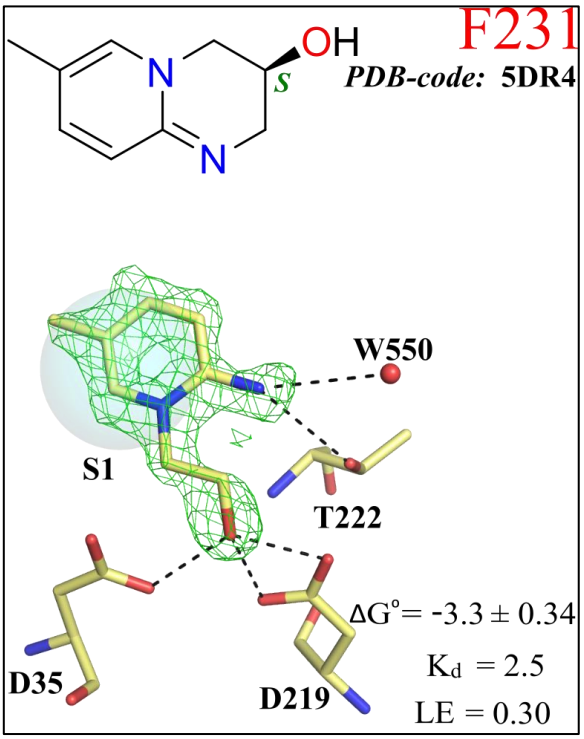


***Fig. 3.10**. Binding mode and affinity data of the alcohol. Representation as in Fig. **3.4**.*

3.4.3 Fragments Binding to the Catalytic Aspartates Mediated by the Catalytic W501.

Even though the larger portion of fragments (24 of 41) binds directly to the catalytic dyad, we also discovered a considerable number of fragments (17 of 41) interacting with the catalytic dyad mediated by the catalytic W501. This water molecule takes a versatile role in binding, as it serves as either an H-bond donor or acceptor with respect to the accommodated fragment.

Primary Amines. In contrast to the above-described fragments that interact directly with the catalytic dyad, seven additional fragments (**F052**, **F058**, **F078**, **F211**, **F285**, **F286**, and **F066**) containing also a primary amino group bind only indirectly to the catalytic dyad. Their contact is mediated by W501, which is appropriately located within H-bond distance to the fragments (2.5−3.0 Å, Fig. 3.11). For all seven fragments, an aromatic ring occupies the S1 pocket forming a π−π-stacking interaction with Phe116. This ring is usually a phenyl, but it can also be a 1,3-dimethylpyrazole (**F058**) or a pyridine bearing a para furan ring (**F066**). In addition, the primary amino group is attached to the central aromatic rings, usually via an aminomethyl group. In the case of **F052**, a 2-aminoethanone moiety binds to the catalytic dyad. This fragment bears a para-bromo phenyl ring.

Fragments **F078** and **F211** are close analogues, as both have an ethyl or methyl ester functionality in the para-position. In the crystal structure of **F211** a second, partially occupied conformer could be modeled into the residual difference electron density (occ. = 31%). The ethyl derivative **F078** could only be assigned up to the methylene group due to insufficiently defined difference electron density for the terminal ethyl group. The primary amines of all seven fragments are likely protonated, donating an H-bond to the carbonyl oxygen of Gly221 and W501. In the structures of **F058**, **F078**, **F211**, **F286**, and **F066**, the conserved W529 or the virtually identical W537 is observed at distances of 2.7−2.8 Å to the fragment's primary amino group.

Furthermore, the meta carboxamide oxygen of **F285** and the para methoxy group of **F286** form H-bonds to W696 at distances of 2.6 and 3.0 Å, respectively. In addition, **F066** (occ. = 70%) interacts with the alternatively modeled W618 (occ. = 30%) that is in a chelating position between the furan oxygen (1.5 Å) and the pyridine nitrogen (1.7 Å).

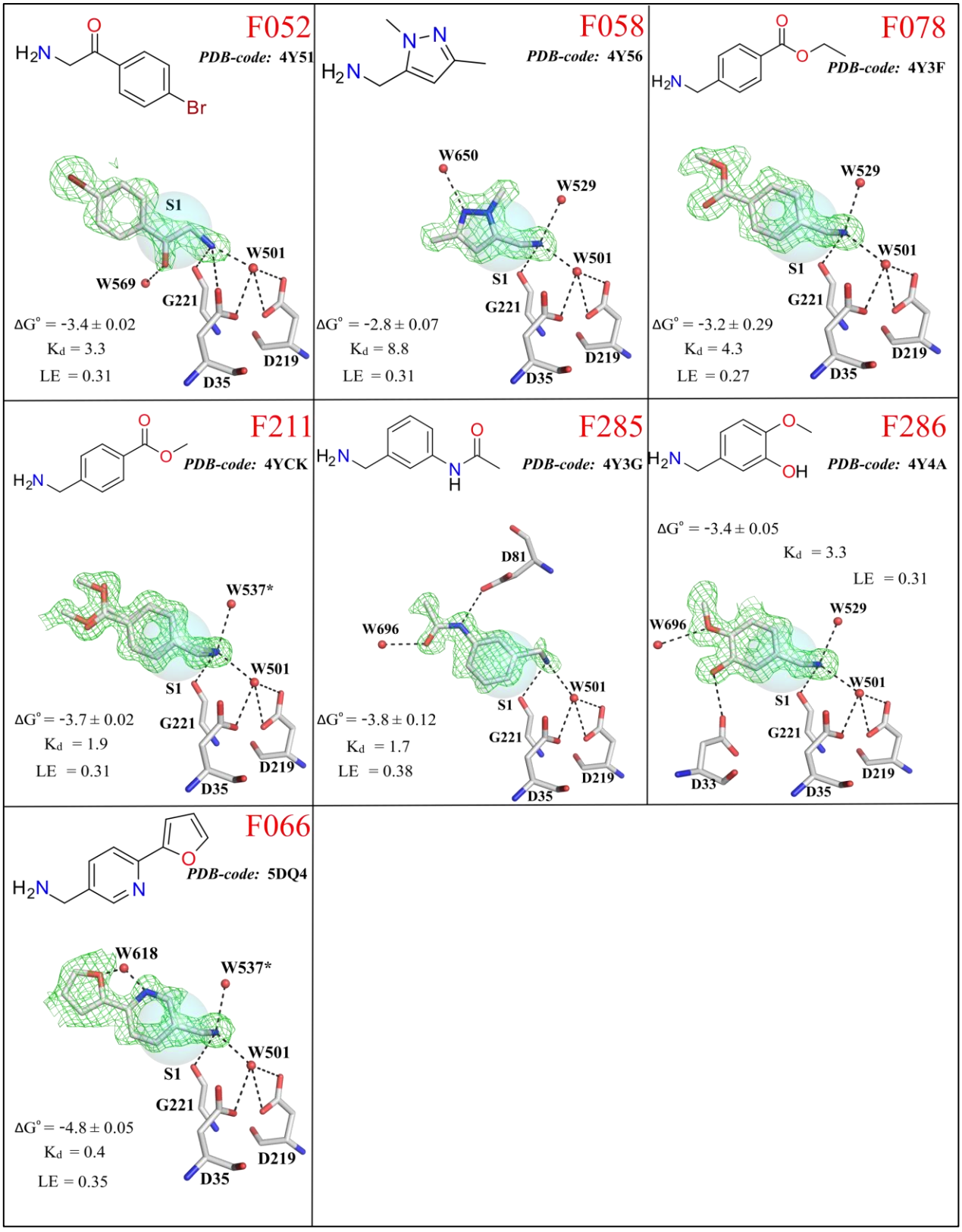


Fig. 3.11*. Binding modes and binding affinity data of the primary amines addressing the catalytic dyad via the catalytic W501. Representation as in Fig.* ***3.4****.*

Pyridine, **Pyrrolidine**, **and Tetrahydrofuran**. Four fragments (**F063**, **F267**, **F272**, and **F291**) address the catalytic water by their pyridine moiety (Fig. 3.12). In all cases, the pyridine rings occupy the S1 pocket with the pyridine nitrogen located in para (**F267** and **F291**), meta (**F063**), or ortho position (**F272**). As suggested by their predicted pK_a values, these pyridine nitrogen atoms are likely deprotonated in aqueous solution at pH 4.6.[36] Furthermore, two molecules of **F291** bind simultaneously to EP. One of them interacts with the catalytic dyad while positioning its benzodioxyl moiety into the S3 pocket (occ. = 100%). In addition, **F063** and **F291** form 2.8 Å H-bonds to Asp81 via their secondary amino group. **F267** performs a similar contact using its amide nitrogen (2.7 Å). Only **F272** accepts an H-bond from Gly80NH with its amide carbonyl oxygen (3.1 Å), as it additionally occupies the S2 pocket with its 1,3,5-trimethylpyrazole ring. Moreover, it forms an H-bond between the amide nitrogen and Thr222OH (3.0 Å). The second molecule of **F291** (occ. = 72%) occupies the S4 and S6 pockets where the aromatic ring system performs a π−π-stacking interaction with Phe291.

Fragments **F112** and **F260** occupy both the S1 pocket with similar poses (Fig. 3.12). However, the pyrrolidine **F112** donates an H-bond to W501 (2.8 Å) with its likely protonated endocyclic nitrogen atom, whereas **F260** accepts an H-bond via its tetrahydrofuran oxygen (2.7 Å). For both fragments, only the *R*-enantiomer is bound. Moreover, the central amide bond of **F260** donates an H-bond to Thr222OH (2.9 Å) and accepts one from Gly80NH (2.5 Å). For **F112**, an additional H-bond is formed between the tetrahedrally coordinated pyrrolidine nitrogen atom and the carbonyl oxygen of Gly221 (2.8 Å). A second molecule of **F112** (occ. = 100%) occupies the S4 and S6 pockets with its benzodioxyl moiety, similar to fragment F291.

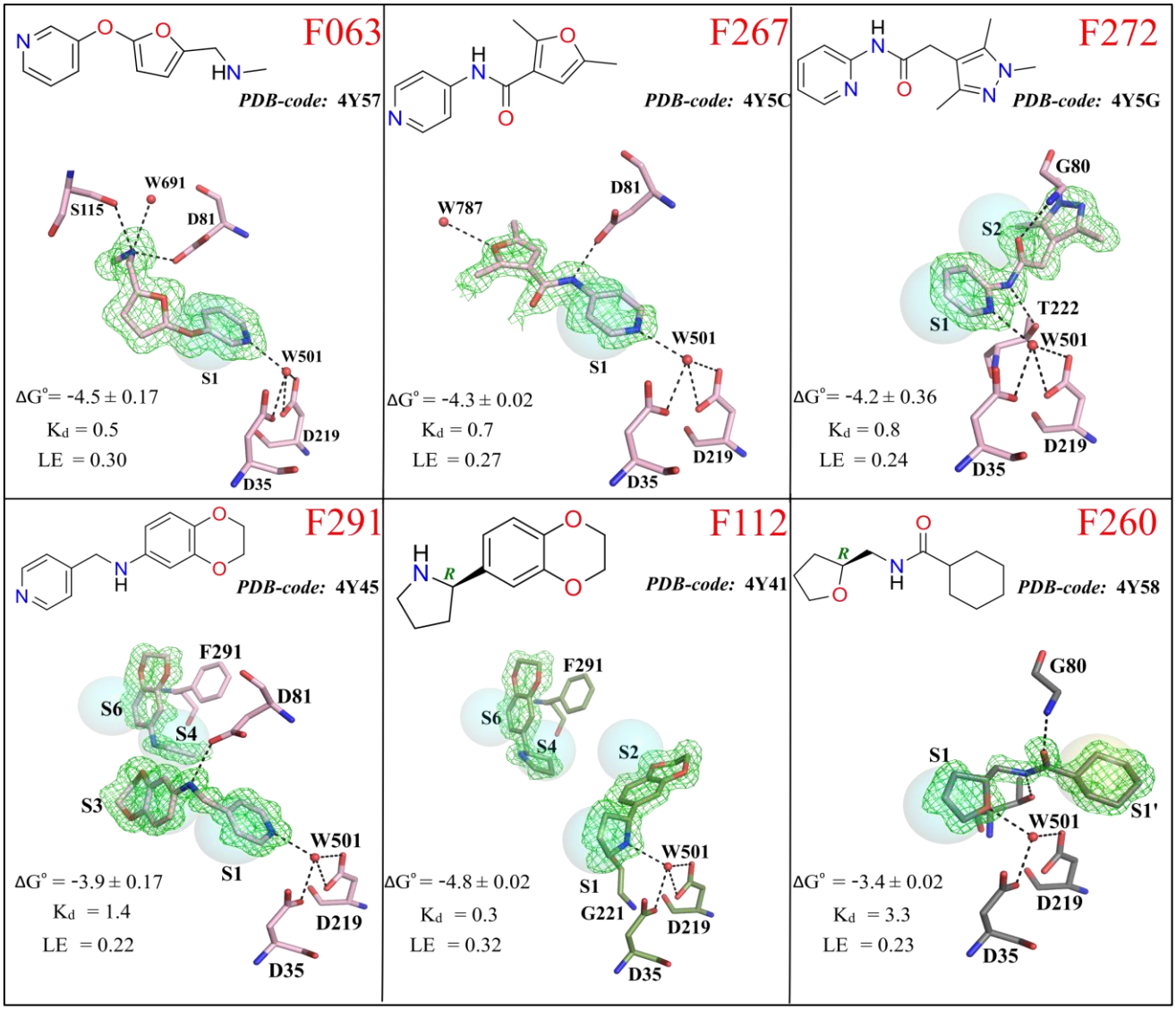


***Fig. 3.12**. Binding modes and binding affinity data of the pyridines, pyrrolidine, and tetrahydrofurane addressing the catalytic dyad via the catalytic W501. Representation as in Fig. **3.4**.*

Oxamide. The oxamide fragment **F048** interacts with the catalytic dyad via one of its carbonyl oxygens, forming a short H-bond to the catalytic W501 (2.2 Å, Fig. 3.13). The second carbonyl oxygen points toward Asp81, accepting an H-bond from the backbone NH (2.9 Å). The fragment places its pyrrolidine moiety into the S1 pocket, while the para substituted pyridine ring occupies the S2 pocket exposing its chlorine substituent to the solvent. An additional interaction is observed between the amide nitrogen next to the pyridine ring and Thr222 (3.0 Å).

Pyrimidine. Two molecules of **F017** bind to EP with full occupancy (Fig. 3.13). In both cases, however, the azepane substituent could only be partially assigned to the electron density, probably caused by the increased conformational flexibility of a seven-membered ring. The likely protonated pyrimidine moiety, which donates a hydrogen bond to the catalytic W501 (2.7 Å), occupies the S1 pocket. The pyrimidine ring of the second molecule forms a π−π-stacking interaction with Phe291.

Thiazolamine. Two molecules of **F171** bind to EP, one interacting with the catalytic W501 (Fig. 3.13) and the second binding remotely on the protein surface. The aminothiazole of F171, which is connected to a methyl imidazole, occupies the S1 pocket where it forms a π−π-stacking interaction with Tyr79. The exocyclic nitrogen forms an H-bond to Asp35 and a second one to the catalytic W501, whereas the endocyclic nitrogen forms an H-bond to W516 (2.7 Å). The methyl imidazole moiety is oriented toward the S3 pocket donating an H-bond from its presumably protonated nitrogen to Ser83 (2.7 Å).

Carboxylic Acid. Remarkably, even the carboxylic acid group of **F266** forms an H-bond to the catalytic W501 (3.3 Å, Fig. 3.13). The fragment is only partially occupied (occ. = 84%) as indicated by the $mF_o - DF_c$ density. The ortho aminopyridine occupies the S1 binding pocket and donates an H-bond to Asp81.

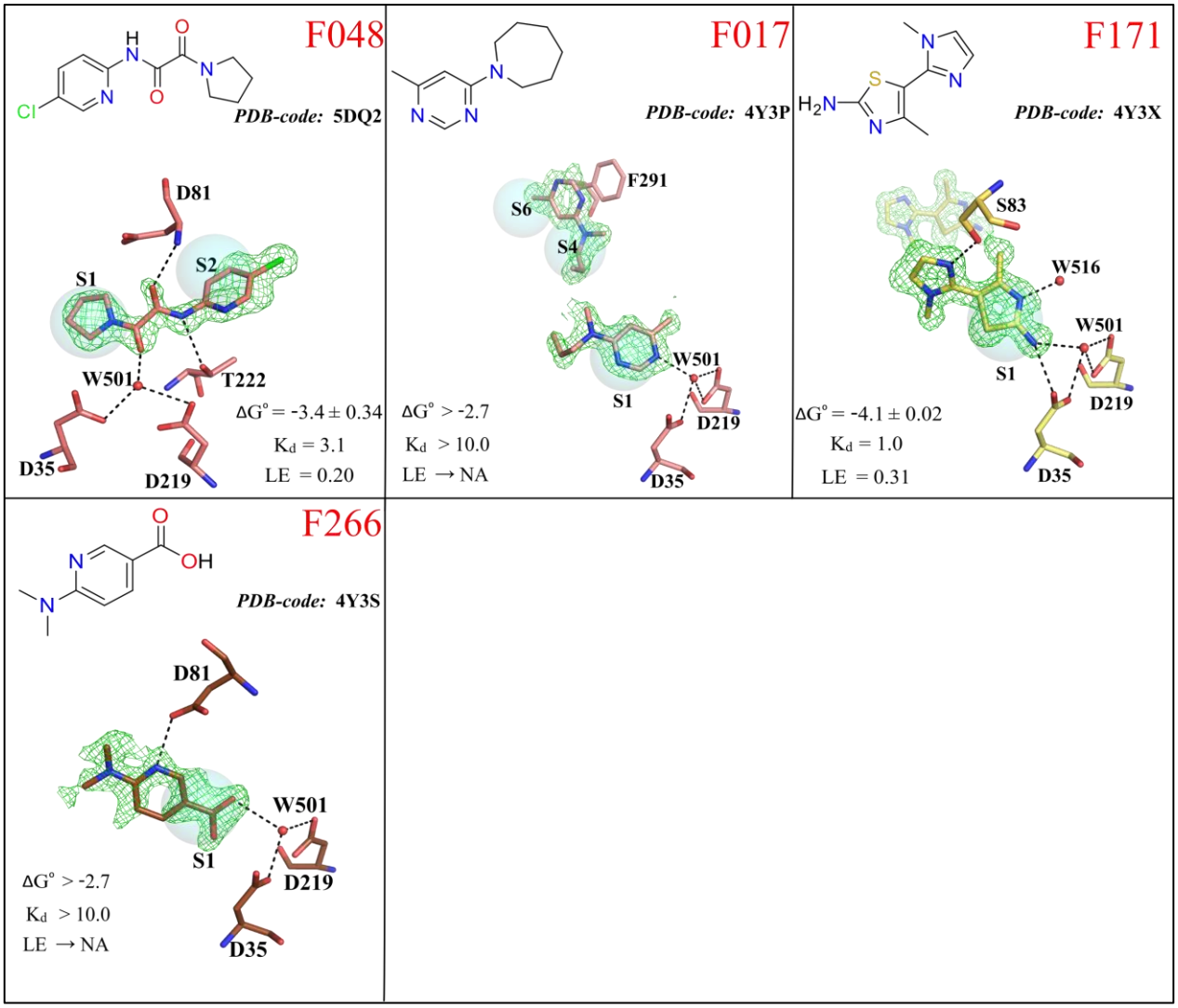


Fig. 3.13*. Binding modes and binding affinity data of the oxamide, pyrimidine, thiazolamine and carboxylic acid addressing the catalytic dyad via the catalytic W501. Representation as in Fig.* ***3.4****.*

3.5 Discussion

In this study, we collected data sets of 361 different EP crystals soaked in solutions containing individual fragments. A total of 41 fragments were found to bind to the catalytic dyad. Twenty-nine additional remote binders are described in an accompanying paper.[23] Of the 41 dyad binders, 24 fragments (59%) bind to the aspartic residues directly, while the interactions of 17 fragments (41%) are mediated by the catalytic water molecule W501 (Fig. 3.1). With regard to their binding regions, 12 (29%) occupy the S1′, 18 (44%) the S1 binding pockets, five (12%) occupy both pockets, and four (10%) are bridging between the S1 and S2 pockets.

Of the remaining two fragments (5%), **F218** is directed toward the S2′ and S3′ pockets (Fig. 3.8), while **F291** occupies the S1 and S3 pockets (Fig. 3.12). In the case of fragments binding multiple times to EP, we only considered those molecules addressing the catalytic dyad in the current analysis.

The most commonly shared chemical functionality of the described fragment hits is a primary amino group. Three fragments (**F005**, **F189**, and **F274**, Fig. 3.5) bind to the catalytic aspartates with this group directly, whereas seven fragments (**F052**, **F058**, **F078**, **F211**, **F285**, **F286**, and **F066**, Fig. 3.11) address the dyad mediated by the catalytic W501. The latter group of fragments clearly differs from the direct binders by the presence of an amino methyl or, in the case of **F052**, an amino ethanone substituent attached to their central aromatic ring. In contrast, only three secondary (**F236**, **F207**, and **F112**) and none of the tertiary amines bind to the catalytic dyad (see also Results section). Since the flexible flap region is positioned next to the catalytic aspartates partially covering the EP binding site, the space for a fragment to be bound is rather limited. An additional restriction is imposed by the spatial proximity of the catalytic aspartates, which explains the highest hit rate for primary amines. In the case of the primary amines, we observed that structural details such as the presence or absence of a chiral center can have unexpected effects on the binding poses of otherwise similar fragments. This is exemplified by the analogs **F052** and **F274** (Fig. 3.14), whose central benzene moieties are substituted by a para bromine or a fluorine atom, respectively, but whose binding poses on first sight are determined by the similar amino ethanone or amino ethanol side chains. While the *S*-enantiomer of **F274** orients its fluorine atom toward the hydrophobic patch formed by Ile300, Ile302, and Ile304 in the S1′ pocket, the achiral amino ethanone **F052** adopts an oppositely oriented binding pose for this fragment. Consequently, the spatially more demanding bromine atom is placed toward the flap region, thus converting a direct dyad binder (**F274**) into one that interacts via the catalytic W501 (**F052**, Fig. 3.14, A). Remarkably, the fluorine atoms of the hydrazide **F103** and the amidine **F216** address the same isoleucine-rich pocket as **F274** (Fig. 3.14, B).

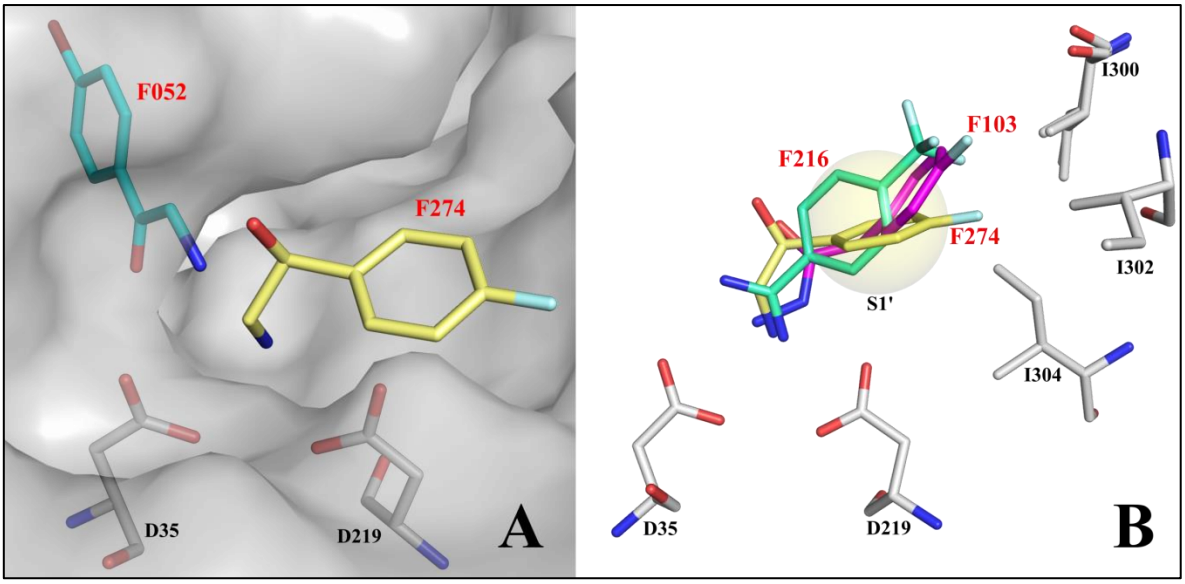


Fig. 3.14. A: *The different orientations of primary amines **F274** (yellow) and **F052** (cyan) are shown. Apart from the different halogen atoms in para position of the phenyl rings in both fragments, the absence of a chiral center in **F052** seems to trigger the conversion from a direct dyad to water-mediated binder. The solvent-excluded surface of apo EP is shown (PDB-code: 4Y5L).* ***B:*** *Superimposition of three fluoro-substituted fragments; hydrazide **F103** (magenta), amidine **F216** (green) and pimary amine **F274** (yellow). All fluorine substituents are oriented into a small cavity formed by three isoleucine residues: Ile300, Ile302, and Ile304. The coordinates of the catalytic dyad, W501, and the hydrophobic S1' pocket composed by the three isoleucines and indicated by a translucent yellow sphere are shown as found in the crystal structure of apo EP (PDB-code: 4Y5L). The surface presentation was omitted for clarity.*

On the basis of this observation, we analyzed whether the bromine substituent of **F052** is actually responsible for the orientation of this fragment. Either steric bulk, unfavorable desolvation penalty, or geometric differences of either an achiral amino ethanone or chiral amino ethanol substituent can serve as an explanation for the observed differences in binding poses. To approach this problem experimentally, we determined the crystal structures of EP in complex with the chiral brominated and chlorinated analogs of **F274** (PDB codes 5IS4 and 5ISJ, Fig. 3.15, A and B) as well as with the achiral fluorinated analog of **F052** (PDB code 5ISK, Fig. 3.15, C). Notably, despite the larger size of the bromine and chlorine atoms, the analogs of **F274** still orient as the fluoro-**F274**, and all three perform an H-bond interaction to Gly80NH via the chiral hydroxyl functionality. In contrast, the fluorinated analog of the bromo-**F052** occupies the S1 pocket and performs π−π stacking with Tyr79 of the flap region. This underscores that the presence or absence of the chiral center in **F274**/**F052** is the determining factor for the surprisingly flipped fragment binding modes.

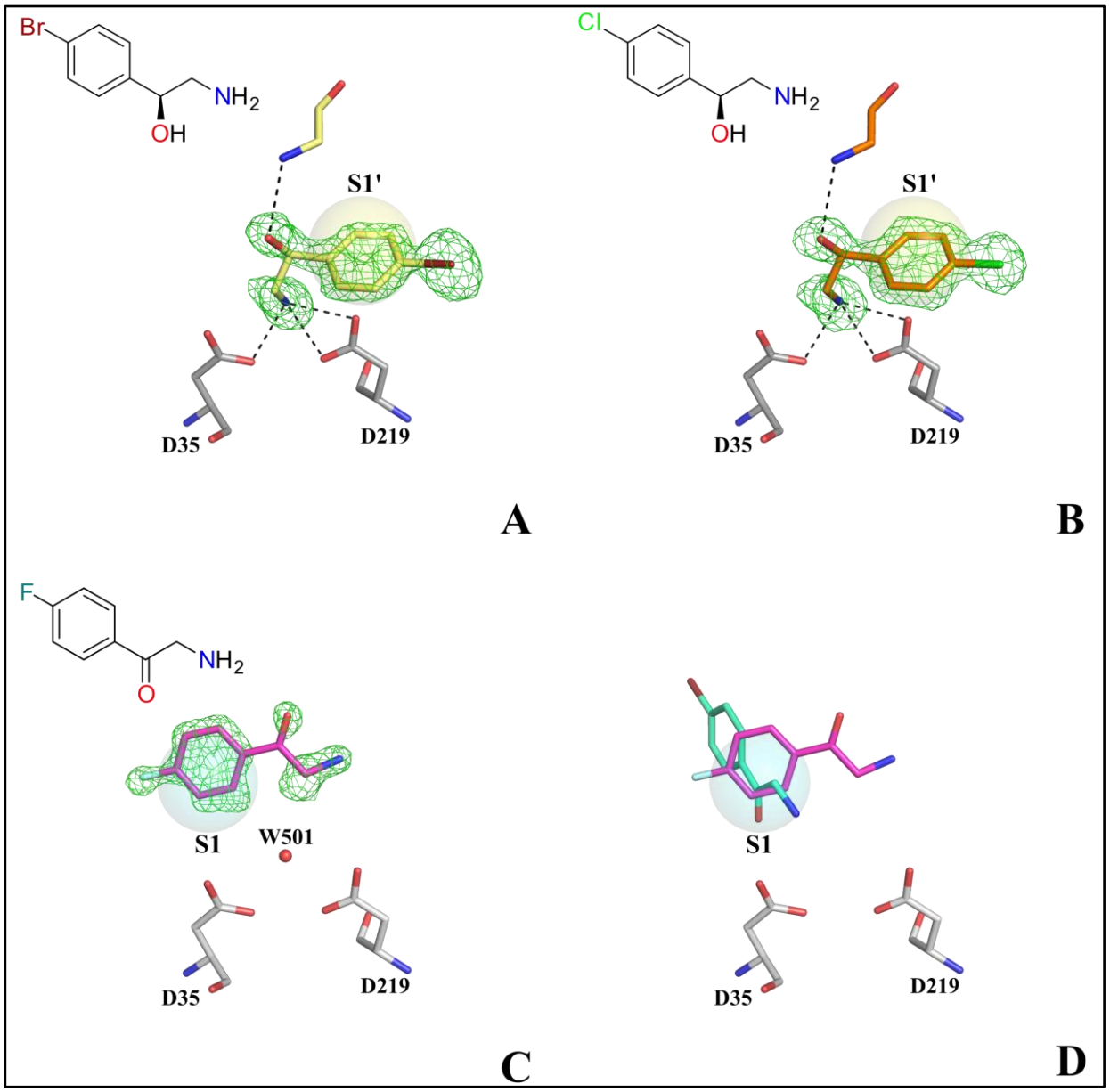


Fig. 3.15. ***A****: Binding mode of the chiral brominated analog of* ***F274*** *(PDB-codes: 5IS4).* ***B****: Binding mode of the chiral chlorinated analog of* ***F274*** *(PDB-codes: 5ISJ).* ***C****: Binding mode of the fluorinated analog of the bromo-****F052*** *(PDB-code: 5ISK).* ***D****: Superimposition of* ***F052*** *(green) and its fluorinated analog (magenta). Representation as in Fig.* ***3.4****.*

To further investigate the structure−activity relationship of this small series, we determined K_d values by isothermal titration calorimetry and calculated ligand efficiencies (LE).[37] We believe that such minute differences cannot be calculated reliably with currently available computational methods. As a result, the K_d values of the halogenated fragments **F052** (K_d = 3.3 mM) and **F274** (K_d = 3.8 mM) do not differ significantly. It should be noted that the K_d value of **F274** was determined for the racemic mixture suggesting an even better affinity of the bound *S*-enantiomer, assuming that the *R*-enantiomer binds with a lower potency to the protein.

While all chiral fragments were screened as racemic mixtures, in five cases only one specific enantiomer was found to bind (Fig. 3.5, 3.7, 3.9, and 3.12). For **F207** both enantiomers were found in the crystal structure (Fig. 3.7). While the *R*-enantiomer addresses the catalytic dyad directly, the *S*-enantiomer occupies the S2 pocket where it is stabilized through van der Waals interactions with the flap region. A third molecule could only be partially assigned to the mF_o-DF_c map next to Phe291; however, the fragment moiety containing the chiral center is missing in the density, thus making its assignment impossible. Obviously, EP is enantioselective with respect to the stereoisomer characterized by the higher binding affinity. This observation suggests that crystallographic screening of racemic fragment mixtures may be reasonable despite the lower concentration of the individual enantiomers. In addition, our results underscore the importance of correct chirality when designing and exploring the structural fragment space by computational approaches in drug discovery.[38]

As discussed above, the protonation state of fragments is key to understand their binding mechanism and structure−activity relationship, particularly if they bind to a charged epitope such as the catalytic center of an aspartic protease. According to pK_a predictions in aqueous solution using the Web service *chemicalize.org*,[36] some fragments, e.g., the pyridines **F162** and **F209** (Fig. 3.9), should be partially protonated in solution but will likely adopt full protonation upon binding, establishing charge-assisted interactions with the deprotonated catalytic Asp35. In contrast, some pyridine fragments that address the catalytic dyad mediated by W501 may do so in either protonated or deprotonated state, which is very difficult to determine with the required certainty. A prediction regarding the protonation state can easily fail, since pK_a values are significantly modulated by the local environment. Nevertheless, it is of critical importance to correctly tailor the protonation states of titratable functional groups of fragments while growing them into follow-up drug candidates. For instance, the charge state is crucial for sufficient bioavailability, as membrane passage often requires uncharged molecules. This highlights that at a first glance negligible design aspects can be of particular importance to correctly tailor not only the affinity but also the physicochemical properties of small molecules.

While primary amines and pyridines along with many other functional groups have already been reported as renin and BACE1 binders,[39–41] here we report for the first time the interaction of hydrazides as putative warheads binding to the catalytic dyad.

Comparison of binding modes of the four hydrazides **F004**, **F081**, **F103**, and **F330** together with **F109** reveals that **F109** is the only fragment oriented in the opposite direction, likely due to its bulkier para diethylamino group (Fig. 3.16). As in the case described further above, the question arises of whether steric restrictions imposed by the volume of the S1 and S1′ pockets provide an explanation for the deviating orientations or whether other features induce the swapped binding modes.

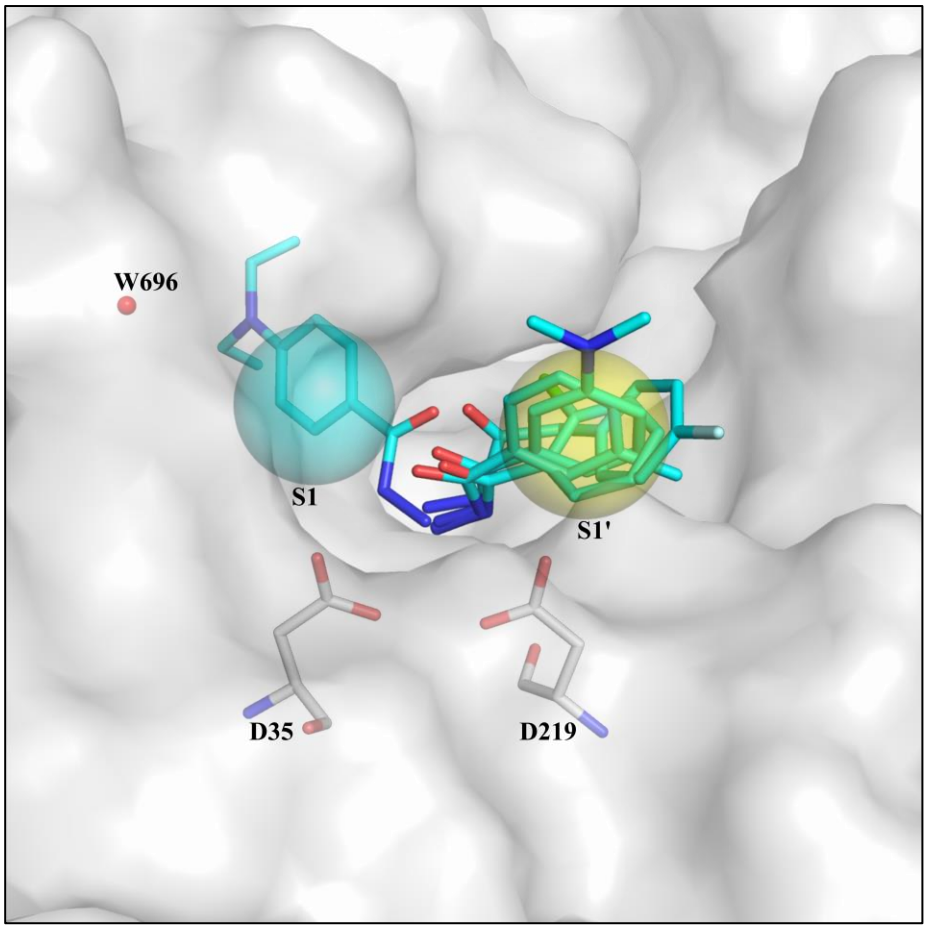


Fig. 3.16*. Superimposition of the five hydrazides **F004**, **F081**, **F103**, **F109**, and **F330**. The coordinates of the catalytic dyad and the solvent-excluded EP surface are shown as found in the crystal structure of apo EP (PDB-code: 4Y5L). The preferentially occupied pockets S1 and S1' are indicated by the blue and yellow spheres, respectively. While four fragments are oriented toward the S1' pocket, only the more potent **F109** occupies the S1 pocket. In the other four complex structures, W696 is present, whereas upon binding of **F109** this water molecule is displaced from the complex.*

To analyze this in more detail, we performed docking experiments (cf. Experimental Section), which revealed that **F109**, even though it carries the largest substituent in the series, should in fact be able to adopt the same binding pose as the other four hydrazides.

This disproved our initial steric crowding hypothesis and shifted our considerations in other directions. By analyzing the crystal structures more carefully, we found that the conserved W696 is displaced by the larger **F109** in the adopted swapped orientation, while in all other EP-hydrazide complexes with the reversed orientation this water molecule is present. This suggests that the fragments with smaller substituents may not be able to displace W696, leading to lower affinity against EP, while attachment of a larger substituent can possibly enhance affinity, most likely for entropic reasons.[42] To validate this hypothesis experimentally, we soaked the slightly smaller dimethylamino analogue of **F109** (Fig. 3.17) and were able to find the same binding mode as for **F109**. By comparison of the K_d values determined by ITC measurements, **F004**, **F081**, and **F330** were characterized as very weak binders (K_d > 10 mM), **F103** with a K_d value of 4.3 mM as slightly stronger binder, while **F109** and the tested dimethylamino analogue are significantly more potent, with K_d values of 1.0 mM and 0.6 mM, respectively. (Fig. 3.4 and Fig. 3.17). Moreover, the previously described biochemical assay showed 89% inhibition for **F109** while none of the remaining terminal hydrazides were detected as inhibitors.[17] This clearly supports our hypothesis and underlines the important influence of water molecules on the binding of fragments, especially with respect to the adopted spatial poses.

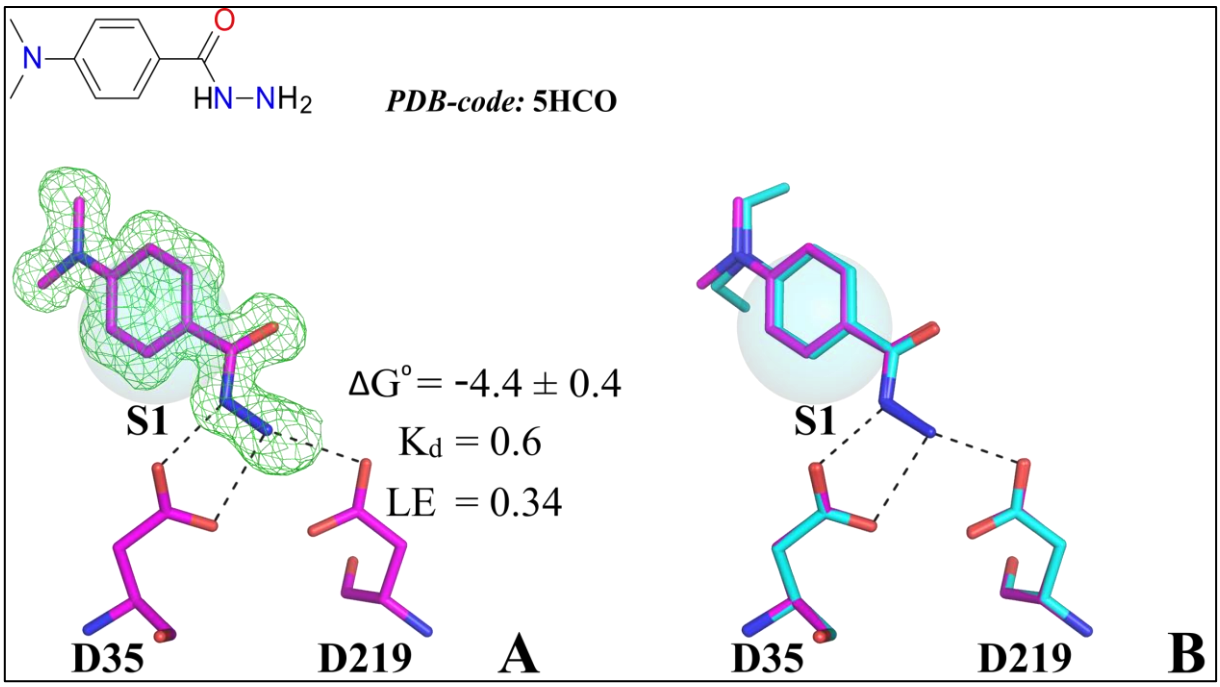


***Fig. 3.17**. **A**: Binding mode and affinity data of the hydrazide analogue of **F109**, directly addressing the catalytic dyad. Representation as in Fig. 3.4. **B**: Overlay of the corresponding crystal structures of F109 and its analogue. The binding mode is almost identical.*

Of particular interest is the in-chain hydrazide **F075** (Fig. 3.4), which occupies the S1′ pocket with its para bromine-substituted phenyl ring and the S1 pocket with a pyrrole moiety. Obviously, addressing both pockets at the same time seems beneficial, as this hydrazide exhibits overall the highest affinity against EP (K_d = 0.1 mM). It favorably bridges the catalytic dyad allowing optimization into both directions of the binding site. Thus, **F075** appears as a highly promising candidate with respect to a fragment-growing strategy.

Another interesting case of swapped binding poses is given for the structurally related amidino fragments.[43] A comparison of the amidino fragments **F125**, **F205**, and **F216** with the isothioureas **F290** and **F306** clearly shows the same orientation relative to Asp35 (Fig. 3.8). Two of the three amidines have affinities in the single digit millimolar range, whereas **F216** and the isothioureas **F290** and **F306** exhibit higher potency in the high micromolar range. All five fragments form a strong bidentate salt bridge to Asp35, requiring however the presence of a charged amidinium group in the ligand (Fig. 3.8). The highly related guanidine derivative **F218** addresses the catalytic dyad in a swapped orientation, preferring Asp219 for the bidentate salt bridge. This binding mode is unique to **F218**, as the quinazoline moiety would clash with the flap in the usually adopted closed conformation. Due to the resulting disorder, the flap residues could not be localized in the electron density. Considering the narrow space available above the catalytic dyad compared to the steric demand of the bulky quinazoline moiety, the adopted deviating orientation of this fragment appears comprehensible. Amidines and guanidines have already been described as warheads for the aspartic protease BACE1,[44] as reported in the field of drug research for neurodegenerative diseases.[45] A phase II/III clinical trial is conducted with the BACE1 inhibitor MK-8931, which originated from a cyclic amidine scaffold used as the key pharmacophore.[46] Also, acylguanidine-like inhibitors were optimized to single digit nanomolar affinity against BACE1.[47−49] Comparing these literature examples with our EP study, we find that the binding mode of the guanidine fragment **F218** to EP is similar to those found in BACE1. However, our isothiourea fragments **F290** and **F306** show a swapped orientation compared to isothioureas detected by a fragment-based NMR screen on BACE1.[50] In BACE1, the isothiourea fragments address Asp289 (corresponding to Asp219 in EP) while **F290** and **F306** bind to Asp35 of EP (corresponding to Asp93 in BACE1). This indicates changes in the pK_a values of the catalytic dyad aspartic residues in the different proteases, leading to deviating fragment-pose orientations. As a result, this might help to address a possible selectivity issue among different members of the aspartic protease family.

Although the use of an amidino-type warhead involves the presence of a charged group in the inhibitor, which will likely lead to problems with oral bioavailability, prodrug strategies have been successfully employed in the past to overcome these difficulties.[51,52]

Dissimilar fragment structures can also adopt similar binding modes. For instance, fragment **F114** (Fig. 3.9) has a similar interaction pattern as observed for the amidine derivatives (**F125**, **F205**, **F216**, **F290**, and **F306**), and it matches with those of the hydrazide derivatives (**F004**, **F075**, **F081**, **F103**, **F109**, and **F330**), although the molecular scaffold providing the contacts is entirely different. This suggests an interesting, not yet reported alternative for bioisosteric replacement, highlighting the value of an exhaustive crystallographic active-site mapping.[53] The examples so far clearly demonstrate that positively charged molecules are favored by EP. However, in a few cases fragments with typical acceptor functional groups can also address the catalytic dyad of the enzyme (Fig. 3.18). For instance, the tetrahydrofuran **F260** and the oxamide **F048** accept a hydrogen bond from the catalytic W501. Quite unexpectedly and reported here for the first time as putative warhead for an aspartic protease is **F266**, which binds to the catalytic W501 via its carboxylic acid functional group. Under the applied soaking conditions this group is likely deprotonated. The adopted charged state at the binding site, however, next to Asp35 and Asp219, is difficult to assess, owing to the crowding of potentially titratable groups. This fragment has an occupancy of only 84% in the crystal structure, which is in accordance with its weak binding affinity (K_d > 10 mM). Although carboxylic acids are present in several drugs (e.g., NSAID aspirin, amoxicillin, or ramipril),[54,55] they can create problems with regard to crossing biological membranes. Nevertheless, for further optimization it is important to recognize that both basic and acidic groups can be incorporated into drug candidates. In close proximity such groups can compensate each other owing to a zwitterionic state and thus improve bioavailability, e.g., as shown for the drug lisinopril.[56] In addition, particularly for carboxylic acid groups, a large variety of prodrug strategies and bioisosters have been reported as successful alternatives.[57,58]

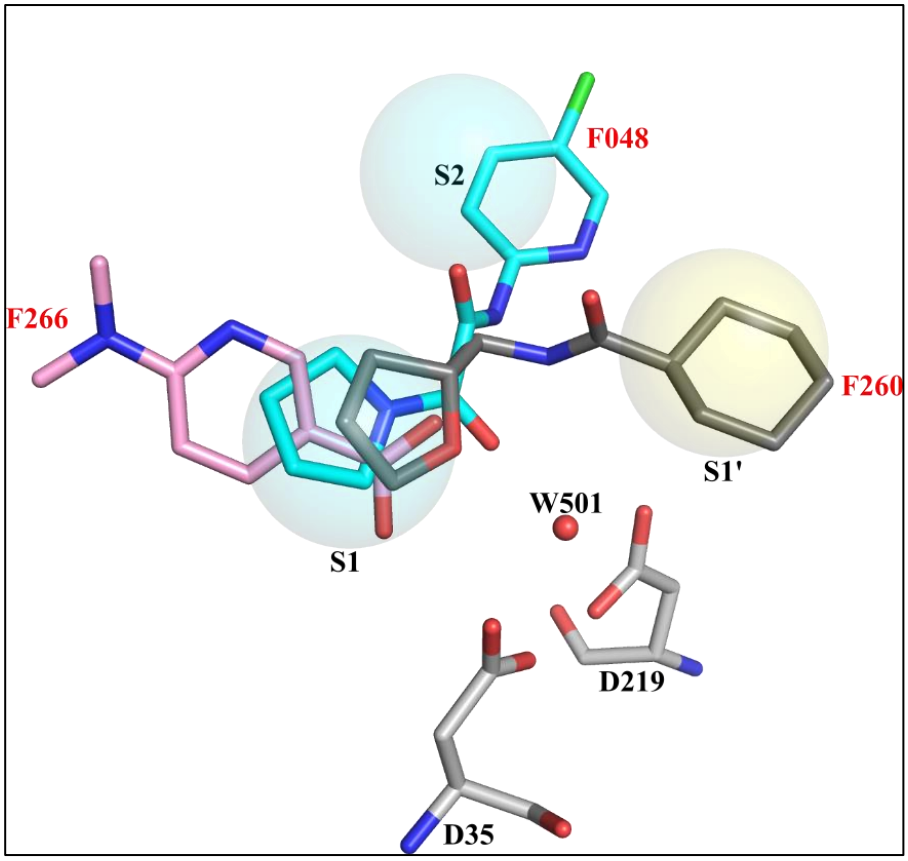


***Fig. 3.18**. Superimposition of the three fragments addressing the catalytic W501 with an acceptor functionality: Carboxylic acid F266 (pink), oxamide F048 (cyan), and tetrahydrofurane F260 (dark gray). The coordinates of the catalytic dyad and W501 are shown as found in the crystal structure of apo EP (PDB-code: 4Y5L). The preferentially occupied pockets S1 and S1' are indicated by the blue and yellow spheres, respectively.*

Our fragment study identified structural motifs as warhead groups to address the catalytic dyad in addition to those that have been previously discovered and incorporated in more potent aspartic protease inhibitors. Still, these fragments provide important information about stereochemistry. For instance, crystal structures of EP with the HIV-protease inhibitor saquinavir or the natural product pepstatin show that both ligands bind to the catalytic dyad with a hydroxyl group, which is also found for the R-configurated fragment **F231**.[59,60] Interestingly, the *R*-enantiomer is clearly preferred by EP, highlighted by its higher affinity and in accordance with a previous study showing higher ligand potency for the pro-(*R*) hydroxyl isomer against renin and EP.[61] The structural information also provides a rationale for the synthetic optimization introducing a gem-diol group, which led to a 10-fold and 90-fold potency increase against renin and EP, respectively. A piperidine derivative, binding to the catalytic aspartates as found in **F207**, has been previously reported as a promising peptidomimetic in the renin research field.[62]

Again, the selectivity of EP for the more potent enantiomer is in agreement with a study where the *R*-configurated piperidine derivative showed 40 times higher affinity against renin than the respective *S*-enantiomer. In another crystallographic fragment screening campaign on BACE1 using cocktails of six fragments, a piperidine moiety was reported to bind to the catalytic dyad.[63] However, the fragment orients toward the opposite direction in BACE1 compared to **F207** in EP. This observation agrees with the abovedescribed swapped orientation found for the isothiourea fragments in BACE1 versus EP.

3.6 Conclusions

The present FBLD study identifies a variety of warheads to address the catalytic dyad of the aspartic protease EP and gives important suggestions with respect to the preferred occupancy of the adjacent specificity pockets S1 and S1′. Some of the observed fragment hits involving functional groups such as piperidines, isothioureas, pyridines, and guanidines have already been reported as potential dyad binders on other aspartic proteases. In addition, we also discovered novel warheads such as hydrazide, carboxylic acid, or pyrazole fragments not yet reported as dyad binders for aspartic proteases. These are interesting candidates for bioiosteric replacements, for instance, when during the drug development phase charged groups have to be avoided or need to be replaced by groups with opposite charge.

The present study focuses on fragments as warheads to address the catalytic dyad. Due to their size, such dyad-binding fragments are in most cases only capable of addressing one of the adjacent S1 or S1′ pockets. The source of ideas on how to grow these fragments into larger and more potent leadlike candidates is thus limited. A few fragments though, such as **F075**, **F114**, **F162**, **F255**, and **F260**, bind to the catalytic dyad and simultaneously occupy both the S1 and S1′ pockets, providing valuable suggestions on how to expand these into larger molecules. To develop such an optimization concept further and even derive initial pharmacophore hypotheses, the information about fragment poses binding to remote pockets is essential. This is discussed in our companion paper.[23]

The majority of discovered dyad binders address either the S1 or the S1′ pocket. Only in the case of the amidino and guanidine derivatives (Fig. 3.8) forming bidentate H-bonds with the dyad steric reasons can explain the swapping of the binding mode.

A comparison of the placement of isothiourea groups in BACE1 and EP indicates surprising differences as to which of the two aspartates is preferred to form a bidentate salt bridge. In the case of the hydrazides (Fig. 3.16), steric bulk cannot explain the swapped orientation of the bound fragments. A docking analysis suggested that both orientations should be feasible. We therefore followed an experimental approach to prove that the displacement of a fairly well-bound water molecule is responsible for the selected binding pose. Indeed, another crystal structure of a related fragment revealed, in addition to the determination of the dissociation constants by ITC, that the swapped orientation is only adopted when the fragments are able to displace the crucial water molecule. For primary amines (Fig. 3.14) we also observed a swapping in the occupation of the S1 versus the S1′ pocket. Here, the slight geometrical change going from a planar carbonyl to a tetrahedral (and thus chiral) hydroxymethylene group influenced the interaction pattern established with the residues next to the catalytic dyad strong enough to alter the observed binding pose.

Without a question, the tendency for EP to accommodate basic, mostly positively charged fragments was evident. Considering the narrow space for binding close to the catalytic dyad, binding of smaller fragments into the binding pocket was preferred over larger molecules, clearly emphasized by the strong preference of EP for primary over secondary and tertiary amines. Nevertheless, fragments with typical acceptor functional groups addressing the catalytic dyad of the protease were also discovered (Fig. 3.18).

Since many of our findings described here would have been impossible without a detailed crystal structure analysis, we strongly advocate consulting crystallography as a primary fragment screening technique. It directly provides valuable insights into the adopted binding modes of the hits, giving important guidance for the subsequent hit-to-lead optimization. This broad and conclusive fragment binding landscape only emerges due to a large number of characterized fragments providing initial ideas about protonation states, preferred pocket occupancies next to the catalytic dyad, and the putative displacement of active-site water molecules whose removal upon ligand binding has a clear contribution to affinity. It is our opinion that these insights are by far more important than the discovery of the most potent or most enthalpic binding hits, as the affinity of a fragment can be strongly affected by the quality of the local water structure.[64,65] During the optimization phase, weak binding fragments can readily be developed into optimal starting points for drug development.[66,67]

From a drug designer's point of view, the wealth of information from a crystallographic fragment-screening campaign with a rich yield is clearly beneficial, as the comprehensive experimental active-site mapping can suggest novel, unexpected binding motifs usually not considered by applying well-established and often no longer scrutinized sets of medicinal-chemistry rules for particular target classes.

3.7 Experimental Section

3.7.1 Endothiapepsin Crystallization and Fragment Soaking.

EP was purified from Suparen (kindly provided by DSM Food Specialties) by exchanging the buffer to 0.1 M sodium acetate at pH 4.6 using a Vivaspin 20 with a molecular weight cutoff at 10 kDa. Protein crystals were grown as already described using the sitting drop vapor diffusion method at 17 °C and equal volumes of mother liquor and protein at a concentration of 150 μM.[17] The reservoir solution consisted of 0.1 M NH_4Ac, 0.1 M NaAc, and 24−30% (w/v) PEG 4000 at pH of 4.6. Protein crystals were observed after 2 days, while the growth period for diffraction-quality crystals was on average 2 weeks. Each fragment was dissolved at 1M stock concentration in DMSO and soaked for 48 h into the empty protein crystals at a final concentration of 90 mM (9% (v/v) DMSO). Furthermore, the soaking drops consisted of 70 mM NH_4Ac and 70 mM NaAc, 16−20% (w/v) PEG 4000, and 19−23% (v/v) glycerol at pH 4.6. Although we encountered solubility problems with some of the fragments during the soaking experiments, it appears impossible to calculate the extent to which such fragments were soluble in the crystallization buffer. The solid deposit, present in these cases, was not removed and thus could have been supportive by constantly readjusting the fragment saturation in solution.

3.7.2 X-ray Diffraction Data Collection.

After 2 days of soaking, the crystals were flash-frozen in liquid nitrogen. The addition of a cryoprotectant was not necessary, since the fragment soaking solution served as such. Synchrotron radiation was an important prerequisite to obtain highly resolved structures of our protein−fragment complexes. The data sets were collected at beamlines 14.1, 14.2, and 14.3 of the BESSY synchrotron in Berlin at a temperature of 100 K. Further details about the data collection and refinement statistics are listed in the Appendix, Table 8.1.

3.7.3 Structure Determination and Refinement.

The diffraction data were indexed, scaled, and merged using *XDS* and *XDSAPP*.[68,69] Occasionally *HKL2000* was used.[70] The crystal structures were determined by molecular replacement using the program *Phaser* from the CCP4 program package or via mediated rigid body refinement using *Phenix*.[71−73] We used the coordinates of a 1.28 Å fragment-bound structure of EP as a reference model (PDB code 3PCW). Refinement was done using *Phenix*, and intermittent model building and improvement were performed with *Coot*.[73,74] All 41 data sets were refined according to predefined criteria. After an initial simulated annealing step, XYZ-coordinates, individual B-factors, and occupancies were refined with five cycles. The further refinement strategy depended on the structural resolution. Isotropic refinement in combination with TLS using eight TLS groups (1−62; 63−81; 82−151; 152−190; 191−204; 205−242; 243−258; 259−330) was performed at resolutions lower than 1.5 Å, while structures with resolution better than 1.5 Å were refined anisotropically. The corresponding refinement strategy was kept if the R-free value was reduced significantly (~0.5%). For ligand building, SMILES codes were first generated with the Web server at *www.molinspiration.com*.[75] The according restraints were generated with GRADE.[76] Water molecules were placed in appropriate difference electron density of at least 3.0 σ, while ligands were modeled at 2.5 σ. Hydrogen atoms were added to the model if the given resolution was appropriate. Furthermore, if a ligand molecule was refined to an occupancy above 85%, full occupancy was assumed. The shown $mF_o - DF_c$ density maps encompassing each fragment were created by removing the fragments from the model. In the case of insufficient different electron density for individual fragment atoms, these atoms were omitted from the model (e.g., **F231**) in order to not introduce model bias. After an additional refinement cycle the depicted maps were generated with *Phenix*. In order not to miss any fragment hit, additional processing of the crystallographic data has been performed by an automated refinement workflow before inspection of the resulting electron density maps.[77] All structural figures were created with *Pymol*.[78]

3.7.4 Generation and Distribution of Binding Site Pockets.

The recognition pockets S1 through S3′, depicted as spheres with a radius of 2 Å in all structural figures, were defined based on three superimposed EP structures in complex with different peptidic inhibitors (PDB codes 2ER6, 4ER4, and 3ER5).[79−81] Pseudo-atoms characterizing the location of each pocket were derived at the center of mass of the three respective P3′ through P3 peptide side chains. However, the P2 histidine of the ligand in 3ER5 was omitted for the placement of the sphere representing the S2 pocket position, since this residue clearly pointed toward the S1′ pocket.

3.7.5 Water Renumbering Procedure.

Unified water numbers were assigned based on a clustering[73,82] of all possible water positions identified via *Phenix* for all refined EP-structures of this project. Moreover, water numbers were consistently increased with increasing distance to the catalytic W501.

3.7.6 Computational Property Predictions.

All 361 fragments were prepared, and their physicochemical properties were calculated within KNIME (KNIME.COM AG).[83] Hydrogens were added to yield the neutral form, and 3D coordinates were generated using CORINA (Molecular Networks GmbH), version 3.491.[84] Subsequently, all solution-state physicochemical properties shown in Table 3.2 were calculated using MOE (Chemical Computing Group Inc.), version 2014.09.[85] While fragment charges at pH 4.6 were assigned by Marvin (ChemAxon LCC),[86] aqueous pK_a values were predicted with the Web service *chemicalize.org*.[36] The FILTER program (OpenEye Scientific Software Inc.), version 2.5.1.4,[87] was used to flag all fragments for the presence or absence of aliphatic amine, primary amine, secondary amine, tertiary amine as well as primary and alkylated aniline functional groups, whereas amidine and/or guanidine functional groups were flagged by the presence of fconv[88] atom type N.gu1, N.gu2, N.mi1, or N.mi2. The resulting properties were then statistically analyzed and compared between X-ray binders and nonbinders using the R package, version 3.1.2.[89]

3.7.7 Docking Experiments.

Docking of F109 to the binding pocket of EP (PDB code 3T6I) was done with the program SEED (Solvation Energy for Exhaustive Docking).[90] SEED exhaustively samples fragment poses and calculates interaction energies using the generalized Born approximation describing solvent effects by continuum electrostatics to account for the effects of bulk water. Hydrogens were placed and their positions minimized with CHARMM using the CHARMm22 force field.[91]

3.7.8 Isothermal Titration Calorimetry (ITC) Measurements.

To determine the thermodynamic signature of the weak binding fragments, displacement titrations were carried out using the nanomolar, strongly enthalpic EP inhibitor SAP114 (PDB code 4LAP). SAP114 was titrated into a solution containing 0.1 M NaAc, 0.25−3.0 mM of the respective fragment, and 3% (v/v) DMSO. As a reference, 0.5 mM SAP114 was titrated into the same solution excluding the fragment.[37] For fragments with K_d values above 10 mM the obtained experimental error in ΔG^0 was $\geq$ 0.5 kcal mol^{-1}. Thus, exact affinity values for these ligands are not provided while defining them as extremely weak binders (K_d > 10 mM).

3.8 References

[1] Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004**, *3*, 660–672.

[2] Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.* **2004**, *47*, 3463–3482.

[3] Davis, B. J.; Erlanson, D. A. Learning from our mistakes: the 'unknown knowns' in fragment screening. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 2844–2852.

[4] Linke, P.; Amaning, K.; Maschberger, M.; Vallee, F.; Steier, V.; Baaske, P.; Duhr, S.; Breitsprecher, D.; Rak, A. An Automated Microscale Thermophoresis Screening Approach for Fragment-Based Lead Discovery. *J. Biomol. Screening* **2016**, *21*, 414–421.

[5] Kranz, J. K.; Schalk-Hihi, C. Protein thermal shifts to identify low molecular weight fragments. *Methods Enzymol.* **2011**, *493*, 277–298.

[6] Cala, O.; Krimm, I. Ligand-Orientation Based Fragment Selection in STD NMR Screening. *J. Med. Chem.* **2015**, *58*, 8739–8742.

[7] Perspicace, S.; Banner, D.; Benz, J.; Muller, F.; Schlatter, D.; Huber, W. Fragment-based screening using surface plasmon resonance technology. *J. Biomol. Screening* **2009**, *14*, 337–349.

[8] Hennig, M.; Ruf, A.; Huber, W. Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. *Top. Curr. Chem.* **2012**, *317*, 115–143.

[9] Wyss, D. F.; Wang, Y.-S.; Eaton, H. L.; Strickland, C.; Voigt, J. H.; Zhu, Z.; Stamford, A. W. Combining NMR and X-ray crystallography in fragment-based drug discovery: discovery of highly potent and selective BACE-1 inhibitors. *Top. Curr. Chem.* **2012**, *317*, 83–114.

[10] Schiebel, J.; Radeva, N.; Koster, H.; Metz, A.; Krotzky, T.; Kuhnert, M.; Diederich, W. E.; Heine, A.; Neumann, L.; Atmanene, C.; Roecklin, D.; Vivat-Hannah, V.; Renaud, J.-P.; Meinecke, R.; Schlinck, N.; Sitte, A.; Popp, F.; Zeeb, M.; Klebe, G. One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* **2015**, *10*, 1511–1521.

[11] Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* **2005**, *48*, 403–413.

[12] Bauman, J. D.; Patel, D.; Baker, S. F.; Vijayan, R. S. K.; Xiang, A.; Parhi, A. K.; Martinez-Sobrido, L.; LaVoie, E. J.; Das, K.; Arnold, E. Crystallographic fragment screening and structure-based optimization yields a new class of influenza endonuclease inhibitors. *ACS Chem. Biol.* **2013**, *8*, 2501–2508.

[13] Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of secondary binding sites in proteins using fragment screening. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15910–15915.

[14] Davies, T. G.; Tickle, I. J. Fragment screening using X-ray crystallography. *Top. Curr. Chem.* **2012**, *317*, 33–59.

[15] Koh, C. Y.; Siddaramaiah, L. K.; Ranade, R. M.; Nguyen, J.; Jian, T.; Zhang, Z.; Gillespie, J. R.; Buckner, F. S.; Verlinde, Christophe L M J; Fan, E.; Hol, W. G. J. A binding hotspot in Trypanosoma cruzi histidyl-tRNA synthetase revealed by fragment-based crystallographic cocktail screens. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 1684–1698.

[16] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug discovery today* **2003**, *8*, 876–877.

[17] Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *J. Med. Chem.* **2011**, *54*, 7784–7796.

[18] Rawlings, N. D.; Waller, M.; Barrett, A. J.; Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **2014**, *42*, D503-509.

[19] Cooper, J. B. Aspartic proteinases in disease: a structural perspective. *Curr. Drug Targets* **2002**, *3*, 155–173.

[20] Geschwindner, S.; Olsson, L.-L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; Beer, T. de; Folmer, R. H. A. Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J. Med. Chem.* **2007**, *50*, 5903–5911.

[21] Cooper, J.; Quail, W.; Frazao, C.; Foundling, S. I.; Blundell, T. L.; Humblet, C.; Lunney, E. A.; Lowther, W. T.; Dunn, B. M. X-ray crystallographic analysis of inhibition of endothiapepsin by cyclohexyl renin inhibitors. *Biochem.* **1992**, *31*, 8142–8150.

[22] Kuhnert, M.; Koster, H.; Bartholomaus, R.; Park, A. Y.; Shahim, A.; Heine, A.; Steuber, H.; Klebe, G.; Diederich, W. E. Tracing binding modes in hit-to-lead optimization: chameleon-like poses of aspartic protease inhibitors. *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 2849–2853.

[23] Radeva, N.; Krimmer, S. G.; Stieler, M.; Fu, K.; Wang, X.; Ehrmann, F. R.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Schiebel, J.; Heine, A.; Klebe, G. Experimental Active-Site Mapping by Fragments: Hot Spots Remote from the Catalytic Center of Endothiapepsin. *J. Med. Chem.* **2016**, *59*, 7561–7575.

[24] Pearl, L.; Blundell, T. The active site of aspartic proteinases. *FEBS letters* **1984**, *174*, 96–101.

[25] Tuan, H.-F.; Erskine, P.; Langan, P.; Cooper, J.; Coates, L. Preliminary neutron and ultrahigh-resolution X-ray diffraction studies of the aspartic proteinase endothiapepsin cocrystallized with a gem-diol inhibitor. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst.* **2007**, *63*, 1080–1083.

[26] Coates, L.; Erskine, P. T.; Mall, S.; Gill, R.; Wood, S. P.; Myles, D. A. A.; Cooper, J. B. X-ray, neutron and NMR studies of the catalytic mechanism of aspartic proteinases. *Eur. Biophys. J.* **2006**, *35*, 559–566.

[27] Coates, L.; Tuan, H.-F.; Tomanicek, S.; Kovalevsky, A.; Mustyakimov, M.; Erskine, P.; Cooper, J. The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray diffraction. *J. Am. Chem. Soc.* **2008**, *130*, 7235–7237.

[28] Lin, Y.; Fusek, M.; Lin, X.; Hartsuck, J. A.; Kezdy, F. J.; Tang, J. pH dependence of kinetic parameters of pepsin, rhizopuspepsin, and their active-site hydrogen bond mutants. *J. Biol. Chem.* **1992**, *267*, 18413–18418.

[29] Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol* **1984**, *179*, 125–142.

[30] Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665-667.

[31] Coates, L.; Erskine, P. T.; Wood, S. P.; Myles, D. A.; Cooper, J. B. A neutron Laue diffraction study of endothiapepsin: implications for the aspartic proteinase mechanism. *Biochemistry* **2001**, *40*, 13149–13157.

[32] R Development Core Team. *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, http://www.r-project.org/, **2008**.

[33] Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comp. Sci.* **2001**, *41*, 856–864.

[34] Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 251–264.

[35] Zhao, Y. H.; Abraham, M. H.; Zissimos, A. M. Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds. *J. Org. Chem.* **2003**, *68*, 7368–7373.

[36] Southan, C.; Stracz, A. Extracting and connecting chemical structures from text sources using chemicalize.org. *J. Cheminf.* **2013**, 5–20.

[37] Schiebel, J.; Radeva, N.; Krimmer, S. G.; Wang, X.; Stieler, M.; Ehrmann, F. R.; Fu, K.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Heine, A.; Klebe, G. Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chem. Biol.* **2016**, *11*, 1693–1701.

[38] Fink, T.; Reymond, J.-L. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *J. Chem. Inf. Model.* **2007**, *47*, 342–353.

[39] Yuan, J.; Venkatraman, S.; Zheng, Y.; McKeever, B. M.; Dillard, L. W.; Singh, S. B. Structure-based design of beta-site APP cleaving enzyme 1 (BACE1) inhibitors for the treatment of Alzheimer's disease. *J. Med. Chem.* **2013**, *56*, 4156–4180.

[40] Ghosh, A. K.; Osswald, H. L. BACE1 (beta-secretase) inhibitors for the treatment of Alzheimer's disease. *Chem. Soc. Rev.* **2014**, *43*, 6765–6813.

[41] Stamford, A.; Strickland, C. Inhibitors of BACE for treating Alzheimer's disease: a fragment-based drug discovery story. *Curr. Opin. Chem. Biol.* **2013**, *17*, 320–328.

[42] Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discovery* **2015**, *14*, 95–110.

[43] Aguirre, C.; Brink, T. ten; Guichou, J.-F.; Cala, O.; Krimm, I. Comparing binding modes of analogous fragments using NMR in fragment-based drug design: application to PRDX5. *PloS one* **2014**, *9*, e102300.

[44] Fobare, W. F.; Solvibile, W. R.; Robichaud, A. J.; Malamas, M. S.; Manas, E.; Turner, J.; Hu, Y.; Wagner, E.; Chopra, R.; Cowling, R.; Jin, G.; Bard, J. Thiophene substituted acylguanidines as BACE1 inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5353–5356.

[45] Butler, C. R.; Brodney, M. A.; Beck, E. M.; Barreiro, G.; Nolan, C. E.; Pan, F.; Vajdos, F.; Parris, K.; Varghese, A. H.; Helal, C. J.; Lira, R.; Doran, S. D.; Riddell, D. R.; Buzon, L. M.; Dutra, J. K.; Martinez-Alsina, L. A.; Ogilvie, K.; Murray, J. C.; Young, J. M.; Atchison, K.; Robshaw, A.; Gonzales, C.; Wang, J.; Zhang, Y.; O'Neill, B. T. Discovery of a series of efficient, centrally efficacious BACE1 inhibitors through structure-based drug design. *J. Med. Chem.* **2015**, *58*, 2678–2702.

[46] Stamford, A.; Strickland, C. Inhibitors of BACE for treating Alzheimer's disease: a fragment-based drug discovery story. *Curr. Opin. Chem. Biol.* **2013**, *17*, 320–328.

[47] Baxter, E. W.; Conway, K. A.; Kennis, L.; Bischoff, F.; Mercken, M. H.; Winter, H. L. D.; Reynolds, C. H.; Tounge, B. A.; Luo, C.; Scott, M. K.; Huang, Y.; Braeken, M.; Pieters, S. M. A.; Berthelot, D. J. C.; Masure, S.; Bruinzeel, W. D.; Jordan, A. D.; Parker, M. H.; Boyd, R. E.; Qu, J.; Alexander, R. S.; Brenneman, D. E.; Reitz, A. B. 2-Amino-3,4-dihydroquinazolines as inhibitors of BACE-1 (beta-site APP cleaving enzyme): Use of structure based design to convert a micromolar hit into a nanomolar lead. *J. Med. Chem.* **2007**, *50*, 4261–4264.

[48] Gerritz, S. W.; Zhai, W.; Shi, S.; Zhu, S.; Toyn, J. H.; Meredith, J. E.; Iben, L. G.; Burton, C. R.; Albright, C. F.; Good, A. C.; Tebben, A. J.; Muckelbauer, J. K.; Camac, D. M.; Metzler, W.; Cook, L. S.; Padmanabha, R.; Lentz, K. A.; Sofia, M. J.; Poss, M. A.; Macor, J. E.; Thompson, L. A. Acyl guanidine inhibitors of β-secretase (BACE-1): optimization of a micromolar hit to a nanomolar lead via iterative solid- and solution-phase library synthesis. *J. Med. Chem.* **2012**, *55*, 9208–9223.

[49] Zhu, Z.; Sun, Z.-Y.; Ye, Y.; Voigt, J.; Strickland, C.; Smith, E. M.; Cumming, J.; Wang, L.; Wong, J.; Wang, Y.-S.; Wyss, D. F.; Chen, X.; Kuvelkar, R.; Kennedy, M. E.; Favreau, L.; Parker, E.; McKittrick, B. A.; Stamford, A.; Czarniecki, M.; Greenlee, W.; Hunter, J. C. Discovery of cyclic acylguanidines as highly potent and selective beta-site amyloid cleaving enzyme (BACE) inhibitors: Part I--inhibitor design and validation. *J. Med. Chem.* **2010**, *53*, 951–965.

[50] Wang, Y.-S.; Strickland, C.; Voigt, J. H.; Kennedy, M. E.; Beyer, B. M.; Senior, M. M.; Smith, E. M.; Nechuta, T. L.; Madison, V. S.; Czarniecki, M.; McKittrick, B. A.; Stamford, A. W.; Parker, E. M.; Hunter, J. C.; Greenlee, W. J.; Wyss, D. F. Application of fragment-based NMR screening, X-ray crystallography, structure-based design, and focused chemical library design to identify novel microM leads for the development of nM BACE-1 (beta-site APP cleaving enzyme 1) inhibitors. *J. Med. Chem.* **2010**, *53*, 942–950.

[51] Hopfner, R. Ximelagatran (AstraZeneca). *Curr. Opin. Invest. Drugs* **2002**, *3*, 246–251.

[52] Ieko, M. Dabigatran etexilate, a thrombin inhibitor for the prevention of venous thromboembolism and stroke. *Curr. Opin. Invest. Drugs* **2007**, *8*, 758–768.

[53] Thornber, C. W. Isosterism and molecular modification in drug design. *Chem. Soc. Rev.* **1979**, *8*, 563–580.

[54] Bodey, G. P.; Nance, J. Amoxicillin: in vitro and pharmacological studies. *Antimicrob. Agents Chemother.* **1972**, *1*, 358–362.

[55] Schölkens, B. A.; Becker, R. H.; Kaiser, J. Cardiovascular and antihypertensive activities of the novel non-sulfhydryl converting enzyme inhibitor 2-[N-[(S)-1-ethoxycarbonyl-3-phenylpropyl]-L-alanyl]-(1S,3S,5S)-2- azabicyclo[3.3.0]octane-3-carboxylic acid (Hoe 498). *Arzneim.-Forsch.* **1984**, *34*, 1417–1425.

[56] Ajayi, A. A.; Campbell, B. C.; Kelman, A. W.; Howie, C.; Meredith, P. A.; Reid, J. L. Pharmacodynamics and population pharmacokinetics of enalapril and lisinopril. *Int. J. Clin. Pharm. Res.* **1985**, *5*, 419–427.

[57] Nielsen, N. M.; Bundgaard, H. Glycolamide esters as biolabile prodrugs of carboxylic acid agents: synthesis, stability, bioconversion, and physicochemical properties. *J. Pharm. Sci.* **1988**, *77*, 285–298.

[58] Ballatore, C.; Huryn, D. M.; Smith, A. B. Carboxylic acid (bio)isosteres in drug design. *ChemMedChem* **2013**, *8*, 385–395.

[59] Behnen, J.; Köster, H.; Neudert, G.; Craan, T.; Heine, A.; Klebe, G. Experimental and computational active site mapping as a starting point to fragment-based lead discovery. *ChemMedChem* **2012**, *7*, 248–261.

[60] Bailey, D.; Cooper, J. B. A structural comparison of 21 inhibitor complexes of the aspartic proteinase from Endothia parasitica. *Protein Sci.* **1994**, *3*, 2129–2143.

[61] Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Sci.* **1992**, *1*, 322–328.

[62] Powell, N. A.; Clay, E. H.; Holsworth, D. D.; Bryant, J. W.; Ryan, M. J.; Jalaie, M.; Zhang, E.; Edmunds, J. J. Equipotent activity in both enantiomers of a series of ketopiperazine-based renin inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2371–2374.

[63] Murray, C. W.; Callaghan, O.; Chessari, G.; Cleasby, A.; Congreve, M.; Frederickson, M.; Hartshorn, M. J.; McMenamin, R.; Patel, S.; Wallis, N. Application of fragment screening by X-ray crystallography to β-secretase †. *J. Med. Chem.* **2007**, *50*, 1116–1123.

[64] Ruhmann, E.; Betz, M.; Fricke, M.; Heine, A.; Schafer, M.; Klebe, G. Thermodynamic signatures of fragment binding: Validation of direct versus displacement ITC titrations. *Biochim. Biophys. Acta* **2015**, *1850*, 647–656.

[65] Betz, M.; Wulsdorf, T.; Krimmer, S.; Klebe, G. Impact of surface water layers on protein-ligand binding: how well are experimental data reproduced by molecular dynamics simulations in a thermolysin test case. *J. Chem. Inf. Model.* **2016**, *56*, 223–233.

[66] Joseph-McCarthy, D.; Campbell, A. J.; Kern, G.; Moustakas, D. Fragment-based lead discovery and design. *J. Chem. Inf. Model.* **2014**, *54*, 693–704.

[67] Baker, M. Fragment-based lead discovery grows up. *Nat. Rev. Drug Discovery* **2013**, *12*, 5–7.

[68] Krug, M.; Weiss, M. S.; Heinemann, U.; Mueller, U. XDSAPP: A graphical user interface for the convenient processing of diffraction data using XDS. *J. Appl. Crystallogr.* **2012**, *45*, 568–572.

[69] Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125–132.

[70] Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **1997**, 307–326.

[71] McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 458–464.

[72] The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 760–763.

[73] Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213–221.

[74] Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.

[75] Bratislava University. Molinspiration. Molinspiration Cheminformatics [Online] **1986**. http://www.molinspiration.com/ (accessed January 2002).

[76] Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G.. Grade Web Server [Online] **2012**. http://www.globalphasing.com/ (accessed February 2012).

[77] Schiebel, J.; Krimmer, S. G.; Rower, K.; Knorlein, A.; Wang, X.; Park, A. Y.; Stieler, M.; Ehrmann, F. R.; Fu, K.; Radeva, N.; Krug, M.; Huschmann, F. U.; Glockner, S.; Weiss, M. S.; Mueller, U.; Klebe, G.; Heine, A. High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* **2016**, *24*, 1398–1409.

[78] DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14–20.

[79] Cooper, J.; Foundling, S.; Hemmings, A.; Blundell, T.; Jones, D. M.; Hallett, A.; Szelke, M. The structure of a synthetic pepsin inhibitor complexed with endothiapepsin. *Eur. J. Biochem.* **1987**, *169*, 215–221.

[80] Bailey, D.; Cooper, J. B.; Veerapandian, B.; Blundell, T. L.; Atrash, B.; Jones, D. M.; Szelke, M. X-ray-crystallographic studies of complexes of pepstatin A and a statine-containing human renin inhibitor with endothiapepsin. *Biochem. J* **1993**, *289*, 363–371.

[81] Foundling, S. I.; Cooper, J.; Watson, F. E.; Cleasby, A.; Pearl, L. H.; Sibanda, B. L.; Hemmings, A.; Wood, S. P.; Blundell, T. L.; Valler, M. J. High resolution X-ray analyses of renin inhibitor-aspartic proteinase complexes. *Nature* **1987**, *327*, 349–352.

[82] Nguyen, T.-D.; Schmidt, B.; Kwoh, C.-K. SparseHC: A memory-efficient online hierarchical clustering llgorithm. *Procedia Comput. Sci.* **2014**, *29*, 8–19.

[83] Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb C.; Thiel, K. Studies in classification, data analysis, and knowledge organization: KNIME: the konstanz information miner. *Data Anal.*; pp 319–326.

[84] Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J. Chem. Inf. Model.* **1994**, *34*, 1000–1008.

[85] Chemical Computing Group Inc. *Molecular Operating Environment (MOE)*; Montreal, **2010**.

[86] ChemAxon LLC. Marvin. http://www.chemaxon.com/, **2014**.

[87] *FILTER*; OpenEye Scientific Software Inc.: Santa Fe, **2013**.

[88] Neudert, G.; Klebe, G. fconv: Format conversion, manipulation and feature computation of molecular data. *Bioinformatics* **2011**, *27*, 1021–1022.

[89] Dean, C. B.; Nielsen, J. D. Generalized linear mixed models: a review and some extensions. *Lifetime Data Anal.* **2007**, *13*, 497–512.

[90] Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* **1999**, *37*, 88–105.

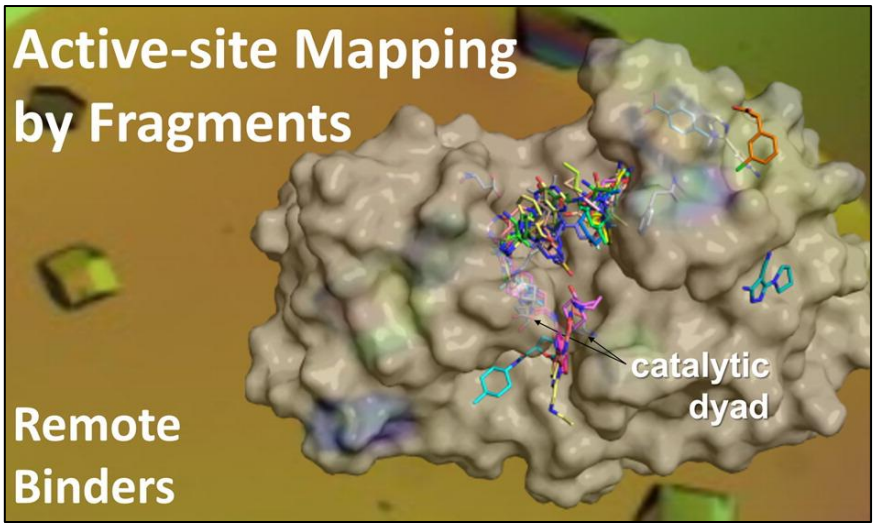
[91] Brooks, B. R.; Brooks, C. L. 3.; Mackerell, A. D., JR; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

4. Experimental Active-Site Mapping by Fragments – Hot Spots Remote from the Catalytic Center of Endothiapepsin

4.1 Introductory Remarks

The following chapter has been published in *J. Med. Chem.* in 2016. The project included the soaking and data collection of over 450 datasets, discussed in the following chapter and in chapter 3. Protein purification, crystallization, and all soaking experiments were performed by the author of this thesis with additional help from Johan Winquist and Ah Young Park. For the data collection and structure refinement, the work of six additional colleagues, Johannes Schiebel, Stefan G. Krimmer, Martin Stieler, Kan Fu, Xiaojie Wang, and Frederik R. Ehrmann was consulted. The data collection was done at Helmholz-Zentrum Berlin für Materialien und Energie, BESSY II with the help of Franziska U. Huschmann, Manfred S. Weiss, and Uwe Müller. All data collection and refinement statistics can be found in the Appendix.

The author of this thesis carried out the refinement of all datasets collected for fragments **F001**–**F070** and performed the data analysis and interpretation for all fragment hits along with drafting the paper.



[*] Reprinted with permission from *J. Med. Chem.* **2016**, *59*, 7561–7575. Copyright 2016 American Chemical Society.

4.2 Abstract

Successful optimization of a given lead scaffold requires thorough binding-site mapping of the target protein particular in regions remote from the catalytic center where high conservation across protein families is given. We screened a 361-entry fragment library for binding to the aspartic protease endothiapepsin by crystallography. This enzyme is frequently used as surrogate for the design of renin and ß-secretase inhibitors. A hit rate of 20% was achieved providing 71 crystal structures. Here, we discuss 45 binding poses of fragments accommodated in pockets remote from the catalytic dyad. Three major hot-spots are discovered in remote binding areas: Asp81, Asp119, and Phe291. Compared to the dyad binders, bulkier fragments occupy these regions. Many of the discovered fragments suggest an optimization concept on how to grow them into larger ligands occupying adjacent binding pockets that will possibly endow them with the desired selectivity for one given member of a protein family.

4.3 Introduction

As enzymes play a key role in the pathogenesis of different diseases, their inhibition is a frequently followed concept for therapeutic intervention.[1] Potent enzyme inhibitors are usually designed starting with a small-molecule lead that mimics the binding of the natural substrate in the enzyme's binding site. However, preferred hot spots of binding for such small molecules have first to be determined and thoroughly analyzed, particularly as this information is of utmost importance to facilitate a subsequent optimization to convert an initial binding hit into a promising lead and finally a drug. Apart from this hot-spot mapping, small-molecules can disclose important aspects of protein function that correlate with induced conformational changes of one or more residues.[2,3] Some screening techniques allow to discriminate between active-site and remote binders. For instance, biochemical assays usually require binding to the catalytic center to produce an assay read-out. In assays such as RDA (reporter displacement assay) and STD-NMR (saturation transfer difference NMR) specific displacement ligands are frequently used to confirm binding to specific parts of a binding pocket.[4–7] Other techniques such as MST (microscale thermophoresis) or the thermal shift assay (TSA) can hardly distinguish binders that address the catalytic center from remote pocket binders.[8,9] Thus, many biophysical screening techniques (named in the following "prescreening techniques") are limited in predicting small-molecule's binding sites.

We therefore performed an elaborate fragment screening to map the extended binding site of the aspartic protease endothiapepsin (EP). In the past, EP has been successfully used as a surrogate for the development of renin and β-secretase inhibitors. Sequence and fold similarity to homologous pepsin-like aspartic proteases along with robust crystallization properties make EP a perfect platform for an exhaustive crystallographic analysis.[10–12]

Out of a well-designed 361-entry fragment library, the overwhelmingly large number of 71 hits was detected. In contrast to our previous study on EP, where the hits of a fluorescence-based biochemical assay were soaked in cocktails to the protein, the intention of this study was to directly screen each fragment in an individual soaking experiment.[13] Of the 71 hits, 70 crystal structures were deposited in the protein databank (PDB) as one fragment requires additional analysis due to a reaction, which likely occurred prior to protein binding. Remarkably, for 14 of the identified fragment hits up to three molecules bind to different pockets of the protein. Considering these fragment copies, in total 86 individual binding poses were determined. Here, we thoroughly analyze 29 fragments binding remote from the catalytic dyad, giving rise to 45 individual binding poses including remotely bound molecules of direct binders. These fragments address hot spots of binding in regions remote from the catalytic center, addressed also by known larger high-affinity EP ligands, demonstrating the importance and utility of fragments for an optimization strategy. In the following Results section, we describe the individual fragment binding poses, interaction patterns including solvent molecules, and order/disorder phenomena in the structures. A comparative analysis also with respect to fragment expansion and lead development follows in the Discussion section.

4.4 Results

To screen the entire library of 361 fragments by crystallography, we performed soaking experiments with individual fragments and EP crystals. The procedure was highly successful as 86 single binding poses of fragments were detected. Of these, 45 molecules were identified addressing remote parts of the EP binding site. Only four fragments bind to the EP surface outside of the substrate-binding site, of which two address the catalytic dyad with a second copy (Fig. 4.1). For the presented study, 39 datasets have been analyzed. These datasets convince through a mean resolution of 1.4 Å, overall completeness of 98.2 %, mean R_{sym} of 7.9 %, and $\langle I/\sigma(I) \rangle$ of 16.9.

The high data quality is highly important to correctly assign fragments in the electron density. The majority of the remote binders (41 out of 45 = 91%) are distributed along the extended EP binding site, four assembled in an area on the protein surface. We split the large binding site into nine pockets, indicated by spheres and labeled according to the Schechter and Berger nomenclature (Fig. 4.1).[14] The most hydrophobic site is the S1' pocket, surrounded by Ile300, Ile302, and Ile304. The S2' pocket is also hydrophobic as it is enclosed by Ile77, Leu133, and Phe194. The S3' pocket is rather hydrophilic exhibiting Ser78, Tyr79, and Gly80 as the polar residues, which are part of the adjacent flap region. Furthermore, the S2 pocket exhibits hydrophilic character as it is enclosed by Thr222, Thr223, and Tyr226. The flap region covers part of the S3 pocket, encompassed by Ile10, Ala16 and Ile122. It merges with the S1 pocket, which is enclosed by Tyr79, Phe116 and Leu125. Accordingly, the S3 and S1 pockets are furnished with amphiphilic character. The S4 and S5 pockets are rather hydrophilic as they are largely exposed to the solvent. Similarly, the S6 pocket is also solvent-exposed but enclosed by hydrophobic amino acids including Leu13, Phe280, Ile283 and Phe291.

The spatial fragment distribution reveals some preferences for certain regions of the binding site. Such sites, besides the catalytic center, are close to Asp119 or Asp81, the latter is part of the flap region. Notably, however, the largest number of fragments is located close to Phe291 (Fig. 4.1). To provide some evidence for preferential binding at particular sites, a thorough understanding of the given structural features and individual attributes of the adopted binding modes of the identified fragment hits is required. Only then, the assembled knowledge can be translated into concepts for a subsequent hit-to-lead optimization. In the following, we will describe the observed binding poses in terms of addressed hot spots and binding pockets.

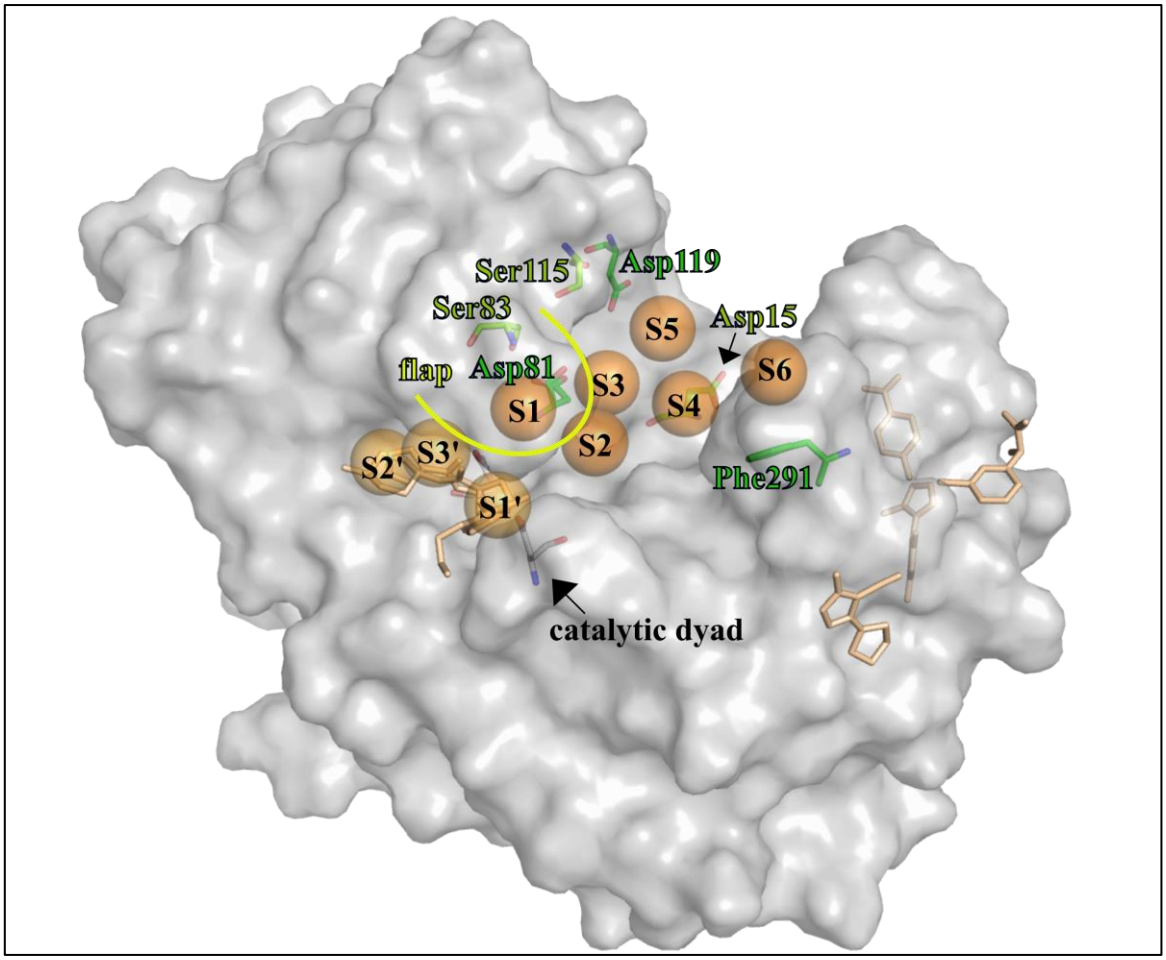


***Fig. 4.1**. EP solvent-excluded surface representation with all defined binding pockets indicated by spheres. (PDB-code: 4Y5L). The determined hot spots of binding (Asp81, Asp119, and Phe291) are shown as green sticks; residues more rarely addressed by fragments are depicted in light green. The flap region is indicated by a yellow line. The four fragments bound on the right-hand side to the EP surface (**F056**, **F114**, **F171**, and **F227**), remote from the catalytic dyad, and the additional four fragments, addressing the S1' and S2' pockets (**F031**, **F131**, **F261**, and **F268**) are shown as beige sticks. The remaining 37 fragment molecules, bound remote from the catalytic dyad were omitted for clarity.*

4.4.1 Fragments addressing Asp81 and the S1 pocket

Asp81 resides, well exposed, on the flexible flap region pointing into the binding site of the *apo* protein. This enables the formation of H-bonds to bound fragments. Thirteen fragments (**F014**, **F125**, **F158**, **F041**, **F323**, **F206**, **F240**, **F073**, **F205**, **F207**, **F224**, **F328**, and **F261**) address Asp81, eleven of which mainly occupy the adjacent S1 binding pocket.

F014 donates a hydrogen bond to the carboxylate oxygen of Asp81 via its methylated thiourea NH group (3.4 Å) and places its *ortho* and *para* halogenated phenyl ring further remote into the S3 pocket (Fig. 4.2, a).

The electron density around the methyl group is ill-defined, suggesting multiple orientations of the latter group. Additionally, **F014** is fixed by Thr223OH, which accepts a hydrogen bond from the second thiourea nitrogen (2.7 Å).

F125 (occ. = 92%), similarly to **F014**, donates an even longer H-bond (3.5 Å) via its terminal amidinopiperidine NH group to the carboxylate oxygen of Asp81 (Fig. 4.2, b). This hydrogen bond deviates by 24° (N-H···O) from ideal geometry due to a slight shift of the aspartate residue compared to its position in the *apo* protein. This shift allows binding of the fragment´s six-membered ring next to the flap region. Furthermore, a fully occupied DMSO molecule is localized in van der Waals distance to **F125** between the S2 and S3 pockets. A second fragment molecule (occ. = 100%) addresses the catalytic Asp35.[15]

F158 likewise occupies two distinct binding sites (Fig. 4.2, c). The first fragment molecule (occ. = 78%) occupies the S1' pocket with its isoxazole moiety, whereas its methylated piperidine moiety is accommodated in the S1 pocket forming an 2.7 Å H-bond between piperidine NH and carboxylate oxygen of Asp81. The piperidine is only partially visible in the electron density. Moreover, the fragment's carbonyl oxygen accepts a hydrogen bond from the backbone NH of Asp81 (3.0 Å), while the fragment's amide donates an H-bond to Thr222OH (2.7 Å). The isopropyl group of the second fragment molecule occupies the S2 pocket, while its remaining part could only be modeled up to the carbonyl oxygen due to an ill-defined electron density.

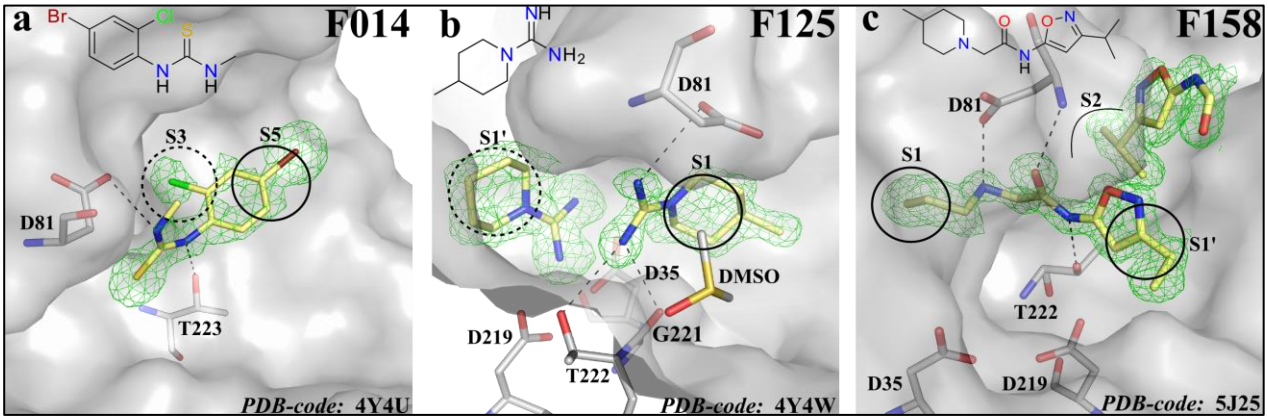


Fig. 4.2. *Fragments, addressing Asp81. In all cases, the $mF_o–DF_c$ electron density around the fragment is shown at 2.5σ, while hydrogen bonds are depicted as dotted lines. Heteroatoms are indicated by the usual color coding, and the carbon atoms of the basic fragment scaffold are colored in yellow. For all experiments, the fragments were used as racemates. The chemical formulas of the bound stereoisomers are displayed in the upper left corner of each image. In addition, the fragment number is shown in the upper right corner and the corresponding PDB code is indicated in the lower right corner. Binding pockets are highlighted as black circles.*

F041, chemotypically similar to **F158**, likewise occupies the S1 pocket with its unsubstituted piperidine moiety and addresses Asp81 and Thr222 identically (Fig. 4.2, d). In contrast to **F158**, only one copy of **F041** is bound, and its piperidine ring is fully visible in the electron density. This might be due to a slightly improved resolution (1.12 vs. 1.24 Å) of the crystal with **F158** or better ordering. The piperonyl moiety of **F041** occupies the S1' pocket.

F323, similar to **F041**, contains a piperonyl moiety, which orients differently as it occupies the S1 instead of the S1' pocket (Fig. 4.2, e). For this fragment, five carbon atoms could not be assigned owing to an ill-defined electron density. Of these unassigned atoms, two carbon atoms belong to the fused piperidine moiety and three are attached methyl groups. A rather short H-bond distance of 2.5 Å is measured from the fragment's endocyclic NH to the carboxylate oxygen of Asp81, whereas the same NH group contacts Ser83OH with a rather long distance of 3.3 Å.

F206 and **F240** both donate H-bonds from their primary amino groups to the carboxylate oxygen of Asp81 (Fig 4.2, f and g). The amino group of **F206** is presumably protonated and donates additional H-bonds to Ser115OH (2.8 Å) and an entrapped buffer acetate (ACT, 2.8 Å). The latter in turn, mediates a further H-bond to the carboxylate oxygen of Asp119, which adopts two alternative conformations. Remarkably, either the acetate or Asp119 must be protonated to establish the observed hydrogen-bonding network, depending on the locally induced pK_a.

F240 establishes, apart from an H-bond to Asp81 (2.7 Å), hydrogen bonds to Ser115OH (2.8 Å), W696 (2.8 Å), and W691 (3.5 Å) via its protonated secondary amino group (Fig. 4.2, g). Furthermore, a glycerol molecule (GOL), which replaces the catalytic W501 and an oligoethyleneglycole (PEG) molecule are found in van der Waals distance to **F240**. The fragment's imidazole moiety forms a π-stacking with Tyr79 in the flap region.

The fragments **F073** and **F205** also donate an H-bond to the carboxylate oxygen of Asp81 and accept one from Ser115OH via their amidino groups (Fig. 4.2, h and i). Moreover, the NH of **F073** accepts an H-bond from Ser83OH (3.1 Å), whereas **F205** donates an additional one of 2.7 Å via its NH group to W696, which mediates the contact to Asp119 (2.9 Å). Furthermore, H-bond contacts between the same NH group, Ser115CO and the second rotamer of Ser115OH, refined to 34% occupancy, are established (3.2 Å and 3.1 Å, resp.).

F205 adopts three binding poses. In two with slightly different orientations, the fragment occupies the S1 pocket (occ. = 45%/55%) and in a third, the molecule interacts directly with the catalytic dyad.

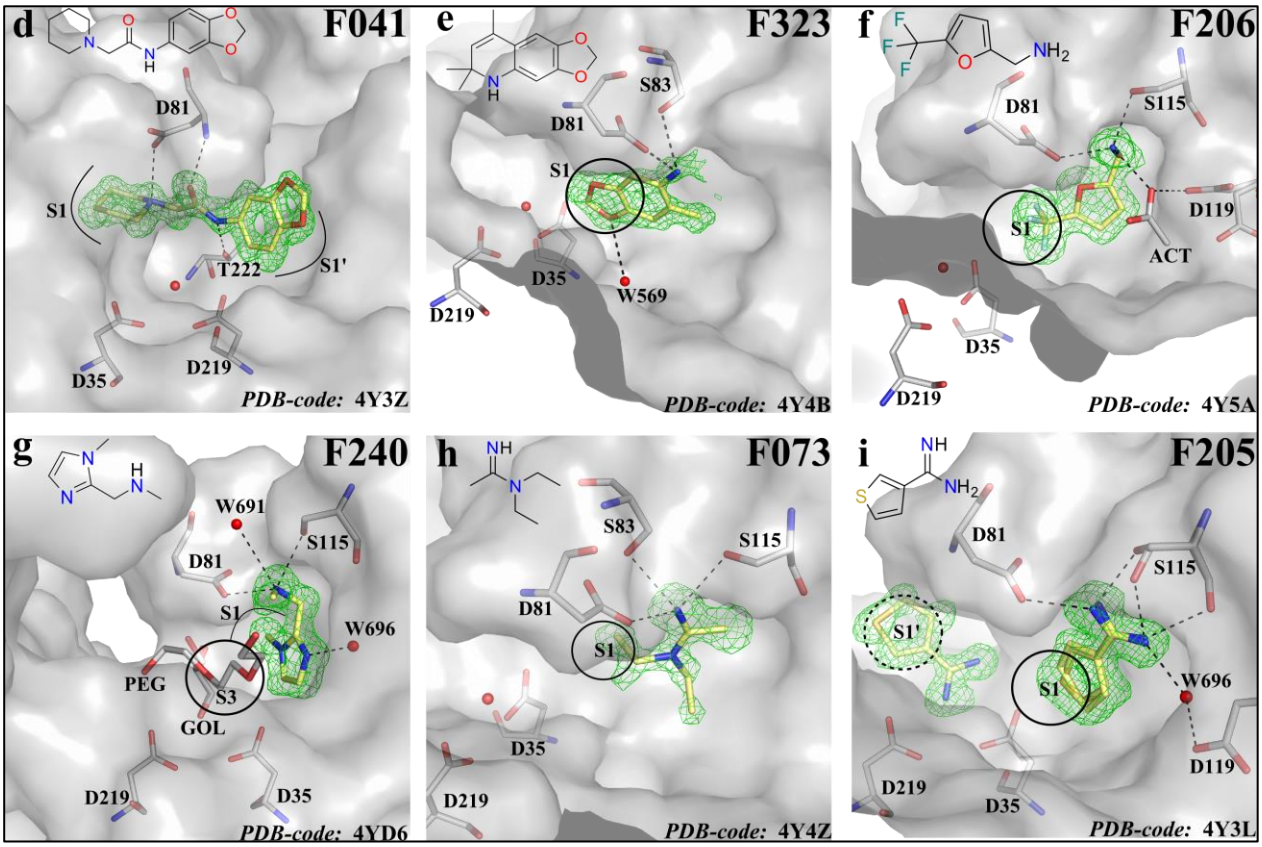


Fig. 4.2. (cont.) *Fragments, addressing Asp81. Representation as in the first* ***Fig. 4.2****.*

F207 occupies three distinct binding sites (Fig. 4.3, a). One fragment molecule binds to the catalytic dyad, a second is found next to Phe291 in the S6 pocket, and a third (occ. = 86%) is located adjacent to Asp81 placing its piperidine-3-yl moiety into the S2 pocket. The latter molecule merely binds via van der Waals contacts to the flap region.

F224, **F328**, and **F261** occupy multiple distinct binding sites (Fig. 4.3, b-d). While **F224** addresses the carboxylate oxygen of Asp81 and Ser115OH with one of its copies, the second copy donates an H-bond to the carboxylate oxygen of Asp119. **F328** and **F261** simultaneously form H-bonds to the carboxylate oxygens of Asp81 and Asp119. Thus, the structural features and binding positions of these three fragments will be described in the following.

4.4.2 Fragments addressing Asp119 and the S3 and S5 pockets.

Asp119 marks an additional hot spot addressed by eight fragments (**F224**, **F328**, **F261**, **F051**, **F321**, **F035**, **F054**, and **F278**), involving adjacent amino acids such as Ser115 and Gly118 into contacts (Fig, 4.3, b-i). Asp119 is localized near the S5 pocket and mostly adopts an alternative conformation compared to the *apo* protein to enable an H-bond with bound fragments.

F224 shows two fully occupied molecules, positioned in mutual van der Waals contact to each other (Fig. 4.3, b). The first addresses the carboxylate oxygen of Asp81 and Ser115OH via its secondary amine (2.8 Å and 2.9 Å, resp.). It occupies the S1 pocket with its methylated pyrrole moiety, which performs a π-stacking with Tyr79 of the flap region. The second molecule addresses the carboxylate oxygen of Asp119 via its indole nitrogen (3.0 Å) positioned in the S3 pocket, where Asp119 adopts an alternative conformation with respect to the *apo* protein to enable this H-bond. The methylated pyrrole ring orients toward the S6 pocket. Interestingly, the central secondary amino group is surrounded by an oligoethyleneglycole (1PE) molecule; this cryoprotectant is frequently entrapped and found at this position (S4 and S6 pockets). However, here it wraps around the fragment in a fashion similar to host-guest interaction found with crown ethers.

F328 occupies two distinct binding sites (Fig. 4.3, c). The first molecule addresses simultaneously the carboxylate oxygens of Asp119 and Asp81 in the S3 and S5 pockets, whereas the second molecule stacks onto Phe291 in the S6 pocket. Asp81 accepts an H-bond from the terminal tertiary amino group of the first molecule, which occupies the S1 pocket (2.7 Å), whereas Asp119 accepts an H-bond from the exocyclic imino nitrogen of the same molecule in the S5 pocket (3.4 Å). Furthermore, the imino nitrogen accepts H-bonds from two neighboring water molecules (W606 and W718). Both fragments are fully occupied and orient with edge-to-face geometry to one another.

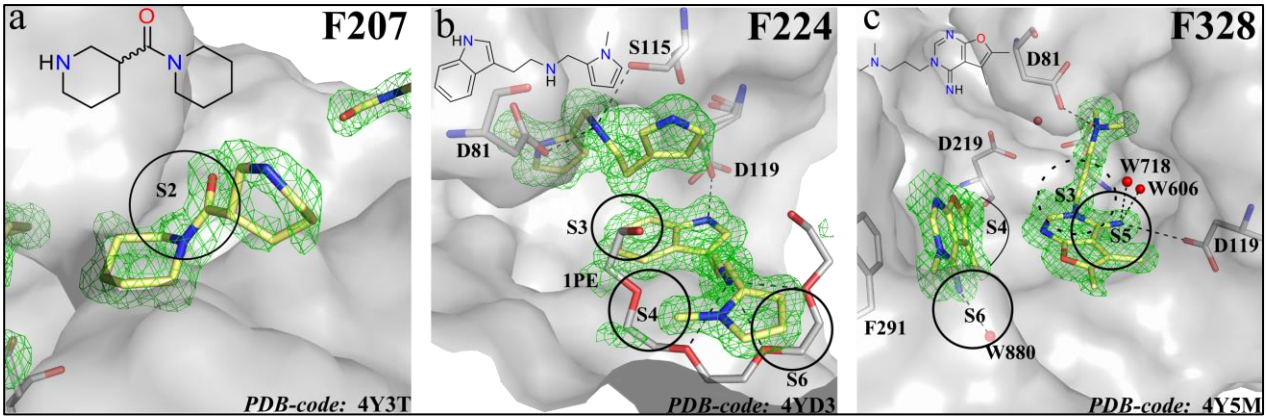


Fig. 4.3. *Fragments, addressing Asp119. Representation as in* ***Fig. 4.2****.*

F261 occupies three distinct binding sites (Fig. 4.3, d). One fully occupied molecule is found in the S3/S5 pockets performing a parallel π-π stacking with a second fragment molecule (occ. = 83%), which is only partially visible in the electron density. The first, fully visible fragment molecule addresses the carboxylate oxygen of Asp119 (3.0 Å) via the 4-N atom of its quinazolinone moiety. To establish this contact, Asp119 adopts an alternative conformation compared to the *apo* state, and either the fragment or Asp119 have to be protonated. Asp81 instead accepts a hydrogen bond from the fragment's amide-like NH group (3.1 Å). Simultaneously, the most likely protonated tertiary in-chain amino nitrogen, located between the S1 and S3 pockets, is recruited for a contact to Asp81 (2.9 Å), while the cyclopropyl moiety orients toward the S1 pocket. The third fragment molecule (occ. = 86%, omitted in the figure for clarity) places its quinazolinone moiety into the S1' pocket while the cyclopropyl group points toward the solvent. The fragment's binding pose could not be fully resolved due to ill-defined electron density.

F051 occupies the S3 pocket (Fig. 4.3, e). It interacts directly with Asp119 through its pyridine-type nitrogen atom (2.7 Å) suggesting a protonated state either on the aspartate or on the pyridine ring. Furthermore, **F051** donates a hydrogen bond with its benzimidazolone moiety to the backbone carbonyl group of Gly221 (3.0 Å) and accepts one from Thr223NH and W629 (3.1 and 2.8 Å, resp.). Furthermore, the pyridine ring interacts with an edge-to-face geometry with Phe116. In addition, GOL is found in van der Waals distance to the fragment, additionally stabilizing its position.

F321 occupies the S1 pocket with its piperonyl moiety (Fig. 4.3, f). The fragment donates an H-bond via its protonated secondary amino group to the carboxylate oxygen of Asp119 (2.9 Å), which adopts an alternative conformation compared to the *apo* protein. A rather short H-bond of 2.5 Å to W655 is formed, which mediates a further interaction to Ser115OH (3.3 Å). In addition, the fragment's piperonyl oxygens accept H-bonds from W511 and W569 in distances of 3.3 Å and 3.2 Å, respectively.

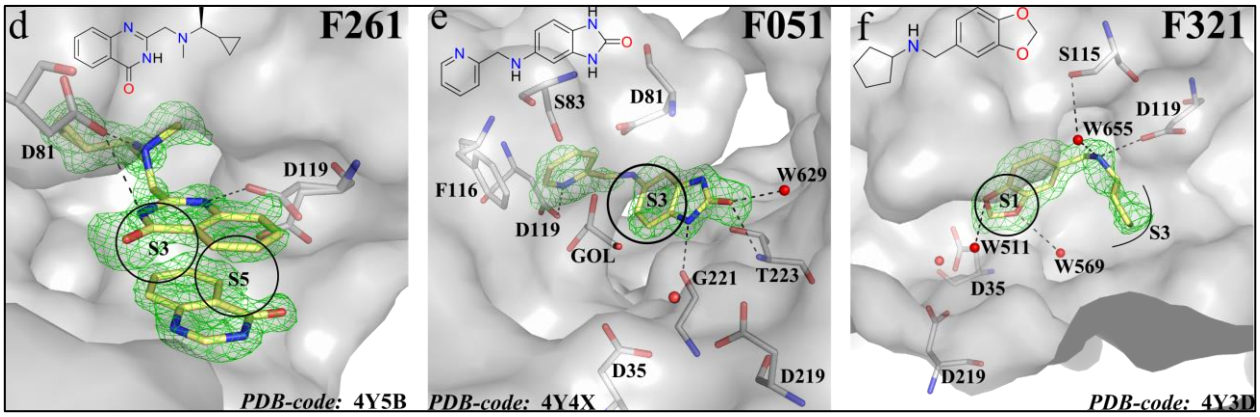


Fig. 4.3. (cont.) *Fragments, addressing Asp119. Representation as in* ***Fig. 4.2****.*

F035 occupies two distinct binding sites in the crystal structure (Fig. 4.3, g). One molecule faces Phe291 in the S6 pocket while the second one places its morpholino moiety into the S3 pocket. The latter molecule addresses Asp119 with its tertiary and primary amino groups, which are supposedly both protonated (2.8 Å and 2.9 Å). Ser115 and Glu118 are also involved in hydrogen bonding to the primary amino group. To establish this contact, Glu118 has to adopt a second conformation compared to the *apo* structure. Both fragment molecules are fully occupied.

F054, similar to **F035**, occupies two distinct binding sites (Fig. 4.3, h). The first molecule interacts directly with the catalytic dyad, while the second one occupies the S3 and S4 pockets donating an H-bond to the carboxylate oxygen of Asp119 via its tetrahydroimidazo triazine moiety (2.9 Å). Additional H-bonds are formed between the tetrahydroimidazo triazine nitrogen of the six-membered ring and W753, which mediates an additional contact to Asp119.

As unique example, **F278** (occ. = 83%) addresses the carboxylate oxygen of Asp119 via a carboxamide oxygen (2.7 Å) (Fig. 4.3, i). The phenyl-cyclopropyl moiety is surrounded by Phe116 and Leu125 in the S1 pocket while the hydroxyl-piperidine substituent occupies the S3 pocket, where also a DMSO molecule from the soaking buffer is entrapped and stabilizes the fragment through van der Waals interactions.

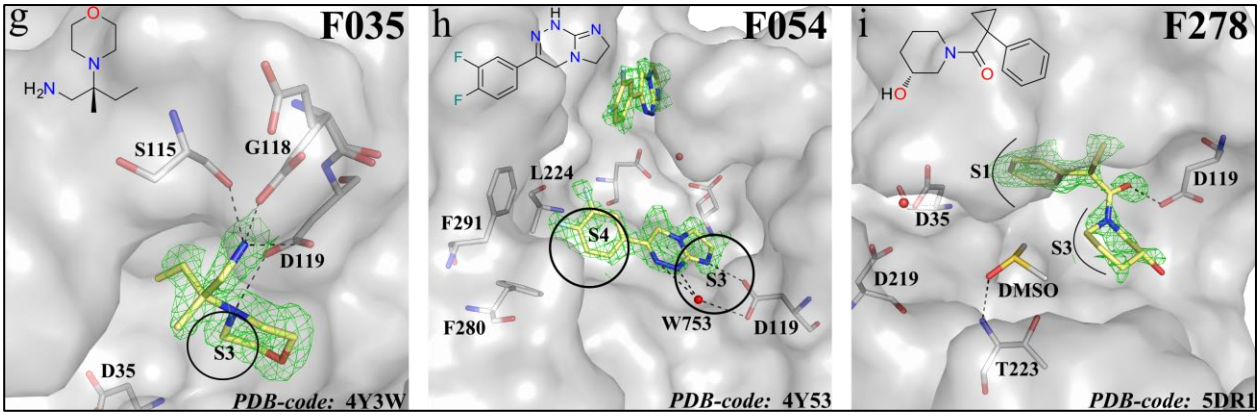


Fig. 4.3. (cont.) *Fragments, addressing Asp119. Representation as in* ***Fig. 4.2****.*

4.4.3 Fragments addressing Phe291 and the S6 pocket.

Phe291 serves as prominent interaction partner in the S6 pocket. In total, sixteen fragments (**F164**, **F034**, **F311**, **F181**, **F337**, **F017**, **F328**, **F333**, **F203**, **F042**, **F305**, **F273**, **F112**, **F291**, **F035**, and **F207**) address this amino acid (Fig. 4.4). Interestingly, in the crystal structure of the *apo* protein an oligoethyleneglycol (PEG) molecule is found at this position.

F164 faces Phe291 with its pyridine ring (Fig. 4.4, a). The fragment's nitrile substituent and the conformationally flexible morpholinopropyl portion could not be placed due to insufficiently defined electron density.

F034 (occ. = 85%) stacks on top of Phe291 with its quinolone-2-one moiety (Fig. 4.4, b). The endocyclic carboxamide group forms H-bonds to W761, while the cyclopropyl moiety points toward the solvent.

In contrast, the cyclopropyl ring of **F311** occupies the space between the S4 and S6 pockets (Fig. 4.4, c). The fragment (occ. = 74%) performs a parallel-displaced π-π stacking with Phe291 using its methoxyphenyl moiety, which is attached in *R*-configuration to a thiazolidin-4-one ring.

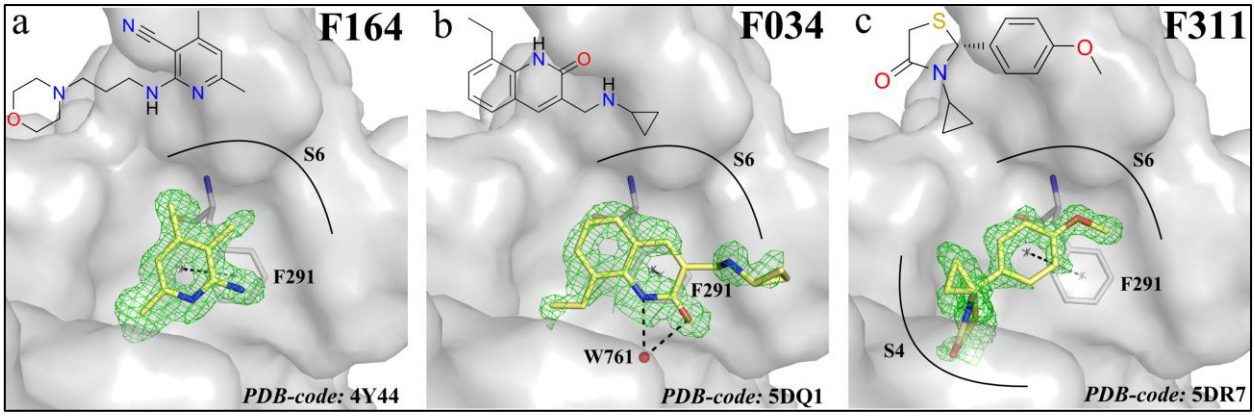


Fig. 4.4. *Fragments, addressing Phe291. Representation as in* ***Fig. 4.2****.*

F181 and **F337** exhibit similar topology and binding modes (Fig. 4.4, d and e). **F181** (occ. = 72%) is substituted by a dimethylpiperidine ring while **F337** (occ. = 100%) bears an azepane ring, attached in *S*-configuration. Although the azepane ring is conformationally more flexible, the density defining the dimethylpiperidine of **F181** is less-well characterized, in agreement with the reduced occupancy of **F181**. Both fragments form a parallel-displaced π-π stacking with Phe291 using their aromatic triazine moieties. Furthermore, Phe280, Pro282 and W1115 are involved in hydrogen bonding via the fragment's exocyclic primary amino functions.

F017 (occ. = 100%) has also a partially visible azepane ring (Fig. 4.4, f). The fragment occupies two distinct binding sites in the crystal structure with one molecule bound to the catalytic W501. The pyrimidine ring of the second molecule stacks on Phe291 in the S6 pocket.

F328, similar to **F017**, occupies two distinct binding sites (Fig. 4.3, c and 4.4, g). The partially visible molecule performs via its heterocycle π-π stacking with Phe291 in the S6 pocket. The spatial orientation of the dimethyl aminopropyl substituent remains unresolved in the crystal structure, likely due to scatter over multiple orientations. Additionally, the exocyclic imino nitrogen donates an H-bond to W880 (2.9 Å), which mediates the contact to the carboxylate oxygen of Asp15 (2.8 Å).

Remarkably, **F333** (occ. = 50%) stacks with its heterocycle on top of Phe291 (Fig. 4.4, h). However, Phe291 adopts a different and so far unique rotamer orientation (occ. = 50%), which is then faced by the fragment. Presumably, the fragment's n-propyl and isopropylsulfanyl substituents reinforce this alternative orientation of Phe291, as at least one substituent would clash with the protein if the usual rotamer of Phe291 would be populated.

F203 (occ. = 64%) has an aromatic pyrazole ring, which is not directly involved in a π-π stacking with Phe291 (Fig. 4.4, i). Instead, the fragment's isopropyl substituent, attached to the sulfonyl spacer, faces Phe291. Furthermore, the primary amino group attached to the pyrazole ring donates an H-bond to the carbonyl oxygen atoms of Pro282 (2.8 Å) and Ile283 (3.3 Å), and accepts one from W1115 (3.4 Å). This example highlights that aromatic ligand portions can be mimicked by an isopropyl group as bioisosteric replacement.

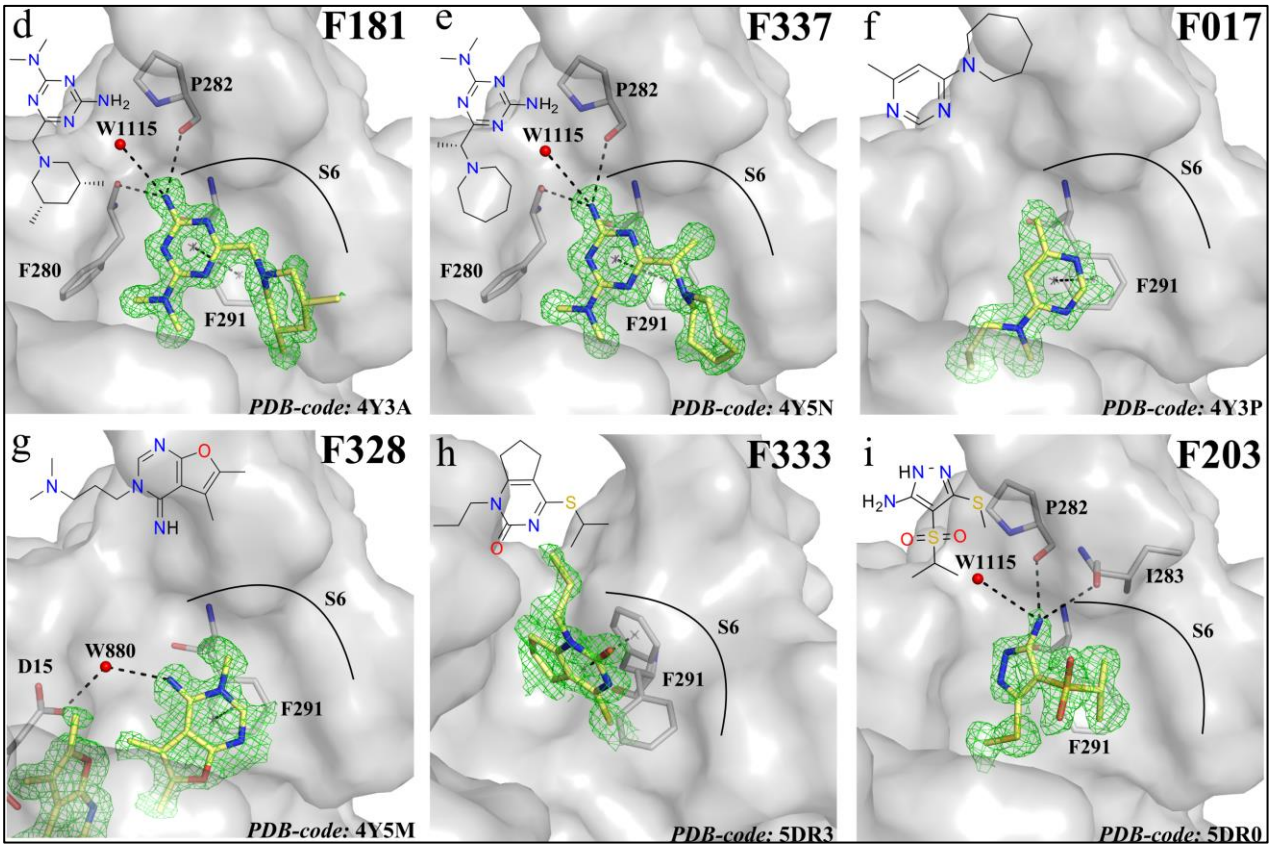


Fig. 4.4. (cont.) *Fragments, addressing Phe291. Representation as in* ***Fig. 4.2****.*

F042 represents a particularly interesting case (Fig. 4.4, j). The fragment stacks on top of Phe291 although the cycloheptatrien ring of **F042** is *per se* non-aromatic. Presumably, it becomes aromatic upon protonation of the imine nitrogen as then the seven-membered ring formally corresponds to a six-π-electron system.

F305, a borinic acid derivative, also stacks on top of Phe291 (Fig. 4.4, k). However, the π-π stacking with Phe291 seems to be performed via the neighboring thiophene ring.

F273 stacks parallel displaced to Phe291 in the S6 pocket (Fig. 4.4, l). **F273** exhibits a dihydroisoquinoline moiety with broken aromaticity at the chiral center, bearing an *N*,*N*-dimethylacetamide substituent. A single H-bond is established to W744 via the fragment's carbonyl oxygen at a distance of 2.6 Å.

F112 and **F291** share similar scaffolds and occupy two distinct binding sites in the respective crystal structures with a second fragment molecule bound to the catalytic dyad (Fig. 4.4, m and n). Both fragments exhibit benzodioxine moieties stacking onto Phe291 and interacting additionally with the carbonylate oxygen of Asp15 via their pyrrolidine or in-chain amino groups.

F035 occupies two and **F207** even three distinct binding sites (Fig. 4.4, o and p and Fig. 4.3, a). The fragment molecules stack on top of Phe291, either via their morpholino (**F035)** or piperidino (**F207**) groups, showing that even aromatic-aliphatic interactions are possible at the Phe291 hot spot.

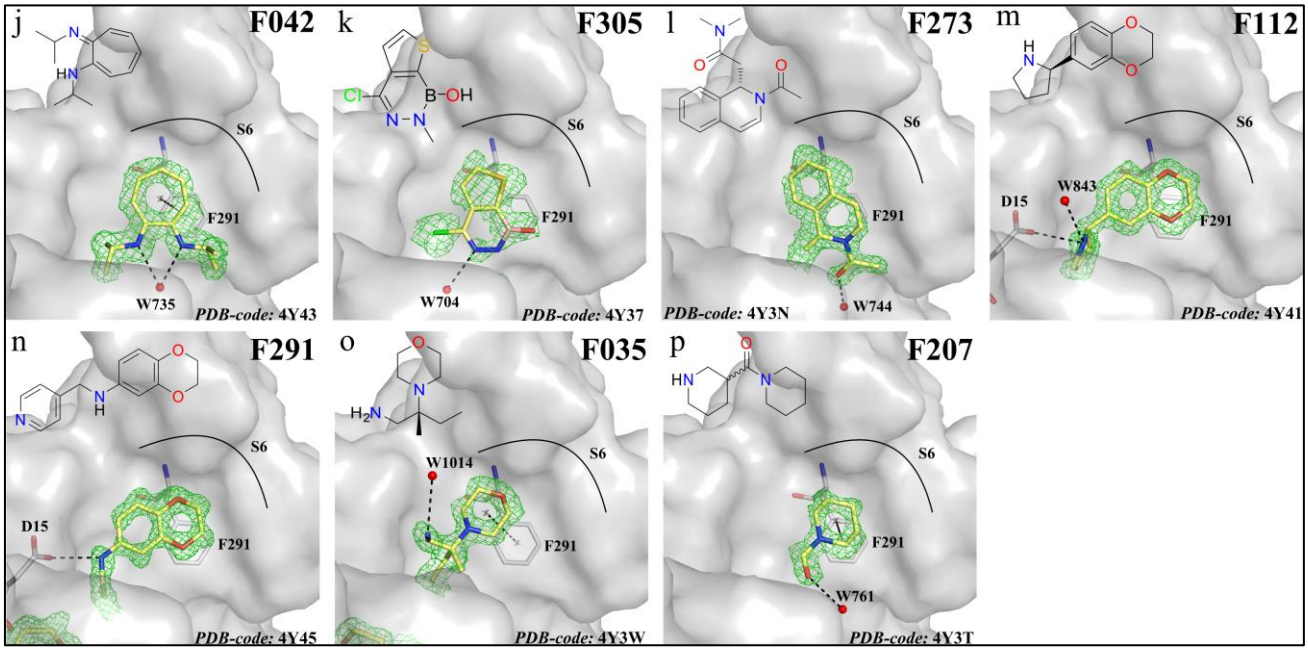


Fig. 4.4. (cont.) *Fragments, addressing Phe291. Representation as in* ***Fig. 4.2****.*

4.4.4 Fragments binding to the S1' and S2' pocket and outside the substrate binding pocket.

F031 occupies the S1' pocket with its phenyl moiety establishing H-bonds with its exocyclic amino group to W515 (2.8 Å) and Thr222OH (3.4 Å) (Fig. 4.5, a). The ethyl ester portion is fully solvent exposed and not defined in the electron density. **F268** and **F131** occupy the S1' pocket using their tetrahydrofurane or trifluoroethyl substituents, respectively (Fig. 4.5, b and c).

Both fragments show a central amide bond, which bridges via H-bonds to Gly80NH and Gly37CO. Notably, the central amide bond in **F131** is placed with reversed orientation between Gly80 and Gly37 compared to **F268**. In addition, both fragments occupy the S2' pocket with the methylated cyclohexyl and isoxazole rings, respectively.

F261 occupies three distinct binding sites (Fig. 4.3, d and Fig. 4.5, d1 and d2). Two molecules occupy the S3 and S5 pockets, while the third one (occ. = 86%) occupies the S1' pocket with its quinazolinone moiety simultaneously exposing its cyclopropyl substituent to the solvent. This substituent is only partially detectable in the electron density.

F114 and **F171** bind with one molecule to the catalytic dyad and a second is placed distal on the protein surface (Fig. 4.5, e and f). **F114** donates an H-bond to Tyr251OH with one of its pyrazolo nitrogen atoms (2.9 Å) and accepts an additional one from W1739 (2.7 Å) via the same nitrogen. The latter water molecule mediates a hydrogen bond to Ser233OH (2.8 Å) and the neighboring nitrogen involves W1398 into an H-bond (2.8 Å). Furthermore, the exocyclic primary amino group donates an H-bond to Ser4OH from a symmetry-related protein molecule in the crystal packing. **F171** forms H-bonds via its exocyclic aminothiazole nitrogen atom, first, to Leu259CO (2.9 Å) and in addition, water-mediated, to Asp276CO (2.8 Å and 2.5 Å). The adjacent Leu259NH addresses the endocyclic aminothiazole nitrogen of **F171** (2.9 Å). Furthermore, the most likely protonated nitrogen atom of the imidazole ring donates an H-bond to W2212 (2.8 Å). The latter water molecule mediates further H-bonds with the carbonyl oxygens of Ala257 (2.7 Å) and Cys255 (2.9 Å).

Two fragments with an acid function bind on the EP surface: **F056**, a *m*-chloro mandelic acid, and **F227**, a *p*-bromomethylene benzoic acid (Fig. 4.5, g and h). **F056** exhibits two water-mediated hydrogen bonds to Ser256OH and Ser289OH, and additionally addresses Ser288OH in a direct contact. **F227** is the only fragment that interacts covalently with EP namely via a nucleophilic substitution reaction with Asp279. As the fragment-binding site is close to the disulfide bridge Cys255-Cys290, possibly the disulfide bridge supports the orientation of the benzoic acid in a way to facilitate the nucleophilic attack of the aspartate's carboxylate oxygen on the fragment's benzyl carbon atom to displace the reactive bromine atom. This example underlines that benzyl halides are quite reactive agents, which can easily attach covalently to polar residues on the surface of proteins.

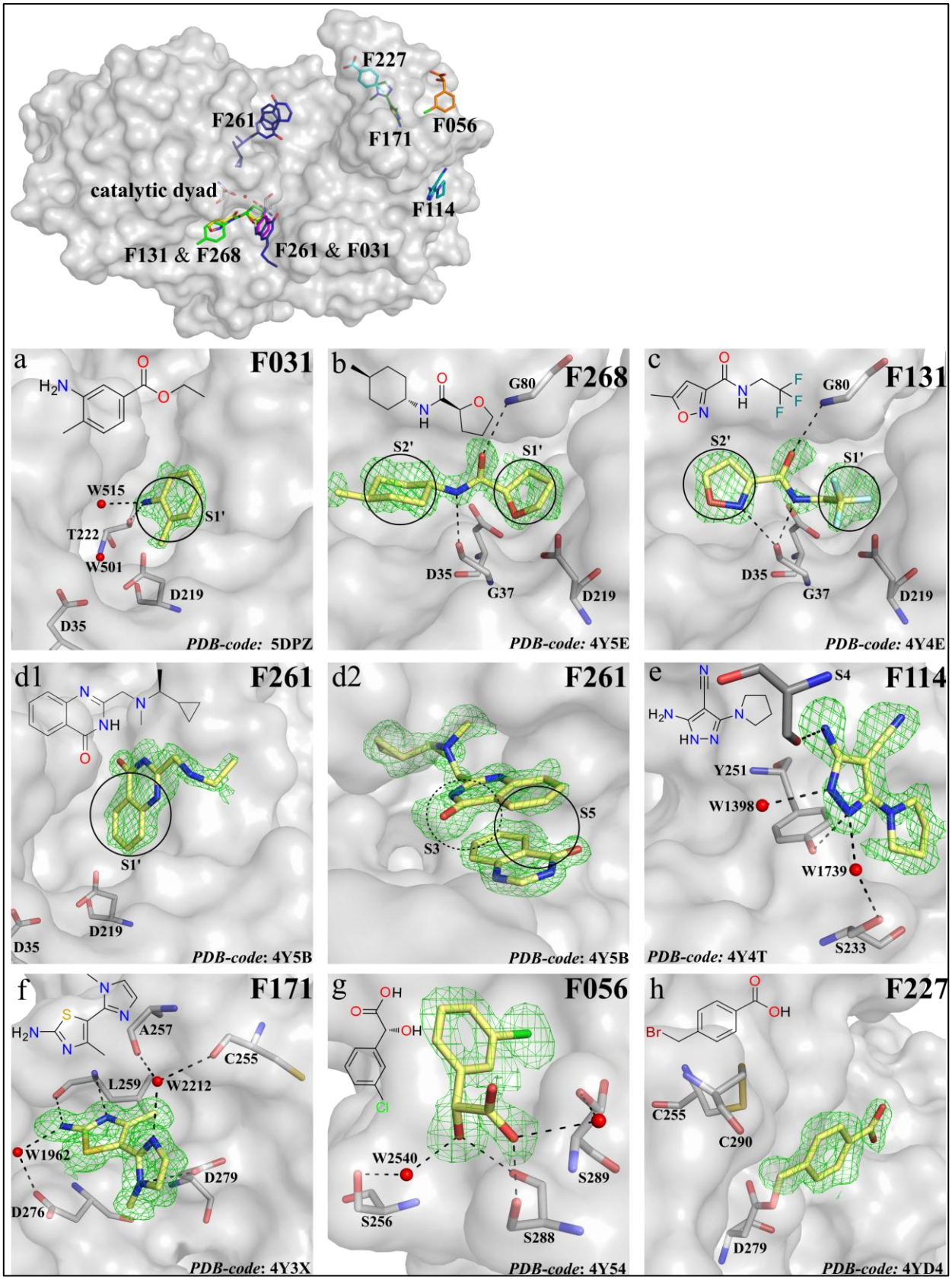


Fig. 4.5. *Fragments bound remote to the catalytic dyad. Representation as in* ***Fig. 4.2****.*

4.5 Discussion

The binding site of EP is quite large (Fig. 4.1) and partly hydrophobic. Hence, it appears very difficult to predict which fragments will bind at which position in this binding site. Although the many aspartic acid residues found in particular in the substrate-recognition site of EP allow the assumption that positively charged fragments would predominantly occupy the binding site, the experimental determination of binding poses are key and can hardly be overrated for any fragment-to-lead optimization procedure. This also becomes apparent from recent studies described in literature. For instance, an X-ray crystallographic screening with more than a hundred fragments on 24 different targets was performed by Astex, and, on average, 2.2 binding sites per target were discovered. The study has a particular emphasis on secondary binding sites. A study with similar focus has been performed by Bauman *et al*., in which an intensive fragment screen on the HIV-1 reverse transcriptase is presented.[16,17] There, halogenated fragments were used to identify hot-spots of binding profiting from the halogens' anomalous scattering signal to detect the bound fragments. The latter group presented in 2013 an X-ray fragment screen against the same target using cocktail experiments, where seven novel protein binding sites were discovered.[18]

Our analysis is based on single soaking experiments with the model protein EP and seeks for the physicochemical properties that distinguish binders from non-binders and, in particular, remote from direct dyad binders, and whether the preferred fragment binding position can be explained by means of these molecular properties. First, the 29 remote binders show, compared to the non-binders, slightly higher mean clogP (o/w) value (1.7 vs. 1.5), and a slightly larger number of Lipinski hydrogen-bond donor atoms (1.4 vs. 1.2, Table 4.1). The difference becomes more significant when comparing the mean number of rotational bonds (3.3 vs. 2.7). The latter value indicates that molecules with higher flexibility can more easily adopt the required binding position in the EP binding site. Notably, while the 29 remote binders do not deviate significantly from the non-binders with respect to their mean molecular weight, the remote binders have, on average, a 32 g/mol higher molecular weight than the catalytic dyad binders (Table 4.2).[15]

Table 4.1 *Comparison of physicochemical properties between the non-binders and 29 remote binders.*

	Non-binders		**Remote binders**		**p-value**[(a)]	
	Range	*Average*	*Range*	*Average*	*Independent samples t-test*	*Mann-Whitney U-test*
No. of non-H atoms	8 - 20	15.0 ± 3.2	9 - 20	15.6 ± 3.3	0.343	0.292
MW (g/mol)	128.2 - 353.3	222.0 ± 43.1	125.2 - 279.6	225.3 ± 40.8	0.680	0.509
clogP(o/w)	-1.25 - 5.39	1.48 ± 0.89	-0.15 - 3.58	1.73 ± 0.99	0.203	0.127
No. of Lipinski donors	0 - 4	1.2 ± 0.7	0 - 3	1.4 ± 0.8	0.299	0.195
No. of Lipinski acceptors	1 - 7	3.7 ± 1.0	2 - 6	3.6 ± 1.2	0.500	0.307
No. of rotatable bonds[(b)]	0 - 10	2.7 ± 1.6	0 - 8	3.3 ± 1.9	0.110	0.070
TPSA (Å^2)	15.3 - 92.7	52.0 ± 13.3	24.4 - 88.8	47.3 ± 15.4	0.119	0.078

[(a)]p-values indicate whether the difference between remote binders and direct binders is significant ($p < 0.05$); please note that for the Mann-Whitney U-test a normal approximation was used
[(b)]Definition according to Oprea considering also ring bonds.[38]

Table 4.2 *Comparison of physicochemical properties between the 41 dyad binders and 29 remote binders.[15]*

	Catalytic dyad binders		**Remote binders**		***p*-value**[(a)]	
	Range	*Average*	*Range*	*Average*	*Independent samples t-test*	*Mann-Whitney U-test*
No. of non-H atoms	8 - 20	13.5 ± 2.9	9 - 20	15.6 ± 3.3	0.008	0.009
MW (g/mol)	122.2 - 288.4	192.6 ± 41.0	125.2 - 279.6	225.3 ± 40.8	0.002	0.001
clogP(o/w)	-0.60 - 3.86	1.20 ± 0.85	-0.15 - 3.58	1.73 ± 0.99	0.022	0.019
No. of Lipinski donors	0 - 4	2.0 ± 1.0	0 - 3	1.4 ± 0.8	0.008	0.019
No. of Lipinski acceptors	2 - 6	3.4 ± 1.0	2 - 6	3.6 ± 1.2	0.412	0.617
No. of rotatable bonds[(b)]	0 - 6	2.6 ± 1.4	0 - 8	3.3 ± 1.9	0.089	0.095
TPSA (Å^2)	29.0 - 99.4	51.0 ± 12.4	24.4 - 88.8	47.3 ± 15.4	0.280	0.283

[(a)]p-values indicate whether the difference between remote binders and direct binders is significant ($p < 0.05$); please note that for the Mann-Whitney U-test a normal approximation was used
[(b)]Definition according to Oprea considering also ring bonds.[38]

In addition, the remote binders show higher mean clogP (o/w) value (1.7 vs. 1.2), and a higher number of rotational bonds (3.3 vs. 2.6) compared to the catalytic dyad binders. Furthermore, considering all of our 361 fragment-library entries, 68 fragments have an amino function (19%). Only two out of all 20 primary aliphatic amines, present in the library, are detected in the EP binding site remote from the catalytic dyad. This ratio gradually increases for secondary (4 out of 23 = 17%) and tertiary aliphatic amines (8 out of 33 = 24%). Along with the information given in Table 4.2, it becomes obvious that fragments with sterically more demanding substituents prefer remote binding sites. As the space above the catalytic dyad is quite limited, only fragments with rather small substituents find a way to occupy the area next to the catalytic aspartates, while the larger fragments simply require more space for binding. For example, a number of bulkier fragments bind to the broad S6 pocket, where, in most cases, fragments with aromatic portions stack on top of the hydrophobic residue Phe291 (Fig. 4.4). In total, sixteen molecules are accommodated in this region (35% of all remote binders). The moieties performing π-stacking with the phenyl ring of Phe291 cover a broad range of diverse chemistry, ranging from aromatic portions (**F311**, **F112**, **F291**) to heteroaromatic rings (**F164**, **F034**, **F181**, **F337**, **F017**, **F328**, and **F333**), thiophene as a phenyl bioisoster (**F305**) to open-chain aliphatic portions (**F203**), cyclic aliphatic fragments (**F273**, **F035**, **F207**) to non-aromatic ring systems, achieving aromaticity upon protonation (**F042**). For example, **F042** and **F017** (Fig. 4.4, f and j) both exhibit a seven-membered non-aromatic ring, namely an azepane and cycloheptatrien, respectively. Nevertheless, at pH 4.6 the imine nitrogen of **F042** is protonated and turns the seven-membered ring into an aromatic tropylium cation, which then performs cation-π stacking interactions with Phe291. At the same time, **F017** addresses Phe291 with its aromatic pyrimidine ring and orients the azepane ring in opposite direction, away from Phe291. Moreover, the attached substituents at the ring used for stacking can take influence on the placement of the fragment molecule and the adopted conformation necessary for the π-stacking interaction.[19] For instance, **F112** and **F291** face Phe291 with their benzodioxine moieties (Fig. 4.4, m and n). Additional H-bonds between the carboxylate oxygen of Asp15 and the fragment's pyrrolidine or in-chain nitrogen atoms, respectively, provoke a slight parallel displacement of the benzodioxine moieties to undergo the π-π stacking with Phe291. Remarkable is also the binding of **F333**, which induces through mutual aromatic stacking, an alternative conformation of Phe291, which exposes its hydrophobic face toward the bound fragment (Fig. 4.4, h).

Moreover, our analysis supports the theory that non-aromatic cyclic polyenes can build favorable non-covalent interactions with aromatic residues of the protein chain as also previously suggested by computational and experimental methods (Fig. 4.4, i).[20,21] Interestingly, an oligoethyleneglycol (PEG) molecule binds either in crystal structures of the *apo* or complexed EP, occupying the S6 pocket next to Phe291, where many hydrophobic fragments bind. In most of the crystal structures with a bound fragment at this position, the PEG molecule is displaced. As a curiosity, one molecule of **F224** penetrates in the semicyclic conformer of the oligomeric PEG molecule and forms within the S6 pocket a kind of host-guest complex. Crown ethers, the fully cyclized analogues of our oligomeric PEG species, are known to host positively charged ions such as potassium in their center. In analogy, the positively charged secondary ammonium ion of **F224** adopts the pivotal position for a rotaxan-type complexation. At the same time, the fragment establishes edge-to-face π-interactions with Phe291. In previous studies, this fragment was detected as hit by five different screening methods and a K_d value of 400 μM was measured via displacement ITC titration.[4,22]

Apart from standard π-π interactions also aliphatic substituents can experience hydrophobic contacts to Phe291. For instance, **F203** represents an interesting example as it orients its isopropyl rest toward Phe291 performing C-H/π interactions (shortest distance = 3.6 Å). Supposedly, as beneficial effect to enhance this interaction, the adjacent acidic sulfonyl group transfers its electron-withdrawing properties onto the isopropyl group, which additionally acidifies the isopropyl substituent rendering its terminal methyl group with a more positive partial charge.[23]

Our study also indicates cation-π interactions, e.g., for **F035** (Fig. 4.4, o), which places its most likely positively charged morpholino moiety close to the π-electron cloud of Phe291 in about 6 Å distance.[24] Remarkably, a similar geometry is found for **F207**, placing its piperidino moiety on top of Phe291 (Fig. 4.4, p). Here, however, the nitrogen is involved in an amide functionality and remains therefore uncharged. The six-membered ring in this complex is closer to the center of the stacked Phe291 (4.4 Å).

The second largest portion of our fragments binds to Asp81 of the flap region. Twelve of the studied fragments occupy the region under the flap (27% of all remote binders). The flap operates as a kind of lid, intermediately opening the binding cleft to give access for substrate binding.

We define this region by the following amino acids: Ser78, Tyr79, Gly80, Asp81, Gly82, Ser83, and Ser84. The high flexibility of the flap is noticeable by the enhanced average B-factors assigned to this region compared to those of the protein main chain atoms in the different crystal structures (Table 4.3). Fragments such as **F240** and **F224** donate strong H-bonds to Asp81 (2.7 Å and 2.8 Å, respectively) and are additionally stabilized by H-bonds to Ser115. These interactions lower the flap's residual flexibility and result in a lower flap-to-main protein B-factor ratio (1.3 and 1.1, respectively). On the other hand, in the crystal structure of **F014** only a weak H-bond between the fragment and Asp81 (3.4 Å) is observed, or no interaction at all is established with the flap for **F035**. Accordingly, higher B-factor ratios result for these complexes (2.1 and 2.9, respectively). This reflects the overall reduced crystal packing order with **F014** and **F035**, and the flexible flap region most likely disturbs this packing. Moreover, while for **F224** an affinity value in the three-digit micromolar range could be measured, significantly lower K_d values were found for **F014** and **F035** (5 mM and 3.2 mM, resp.). However, **F240** was determined as very weak binder with a K_d value greater than 10 mM.[22] This underlines the importance of structural information in order to purposefully grow a basic molecular scaffold using fragment data to optimally fill a certain region of the substrate's binding site without relying solely on affinity values.

***Table 4.3**. Comparison of the mean B-values of the flap region with those from the protein chain.*

Amino acid addressed	Fragment #	average B-factor of the flap region	average B-factor of EP	ratio flap/EP[b]
Asp81	F240	18.2	13.9	1.3
	F158	17.4	12.2	1.4
	F206	15.7	10.8	1.5
	F041	18.6	12.5	1.5
	F073	19.0	12.3	1.5
	F323	22.8	13.5	1.7
	F205	18.9	10.0	1.9
	F125	26.5	12.8	2.1
	F014	38.8	18.6	2.1
Asp119[a]	F224	13.6	12.2	1.1
	F261	19.3	13.8	1.4
	F328	20.4	13.3	1.5
	F278	21.9	13.6	1.6
	F054	20.6	12.0	1.7
	F039	32.9	19.1	1.7
	F051	23.6	11.8	2.0
	F035	52.7	17.9	2.9

[a] *Asp119 is not a part of the flap, but most of the fragments reach this residue and influence its flexibility.*
[b] *The ratio was calculated by dividing the flap's B-factor by the average EP's B-factor including the flap region.*

Considering all twelve fragments bound to Asp81, it becomes obvious that they all share a positively charged nitrogen atom in common, enabling the formation of a direct H-bond to the aspartate. Ten of the twelve fragments additionally occupy the S1 pocket, while **F014** places its sterically more demanding halogenated phenyl ring into the further remote S3 and S5 pockets. Moreover, also one molecule of the triply populated **F207** is in close vicinity to Asp81, but it does not interact directly with this residue, as the distance is too large. In consequence, it prefers to occupy the S2 pocket.

Interestingly, the two fragments **F224** and **F240**, which reduce the mobility of the flap at most, adopt related binding poses, presumably because of their similar topology (methylpyrrole and methylimidazole, resp.) (Fig. 4.6). **F240** and one copy of **F224**, both fill the S1 pocket and interact with the same residues. In both crystal structures, GOL and PEG bind in van der Waals distances to the fragments, and the additives further stabilize the position of the fragments. Interestingly, the GOL molecules in the binding site of both crystal structures bind between the catalytic aspartates, obviously displacing the catalytic water molecule found in the *apo* crystal structure at this position.

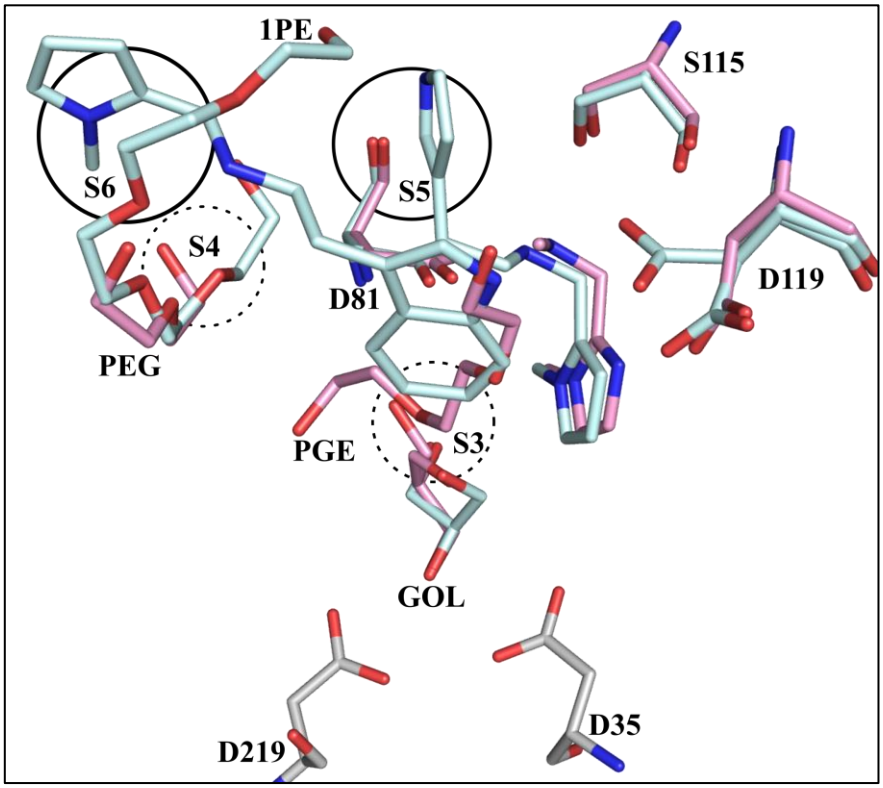


***Fig. 4.6**. Superimposition of **F224** (cyan) and **F240** (pink).*

Other chemical groups that are shared by some of the fragments do not bind, however, in the same region of the large EP binding site. While **F323** and **F321** orient their piperonyl substituents identically, **F041** places this group with totally different orientation into the S1' pocket (Fig. 4.7). In the latter case, the piperidino substituent likely determines the binding pose and is accommodated in the S1 pocket to establish a 2.9 Å H-bond to Asp81. This example highlights the challenge to selectively target a binding site as big as that of EP.

The third hot spot in the EP binding site to which eight fragment hits are bound (18% of all remote binders) is Asp119 (Fig. 4.3). All eight fragments occupy the S3 pocket. Three are found with only one molecule in the respective crystal structures (**F051**, **F321**, and **F278**), whereas five occupy two distinct binding sites (**F224**, **F328**, **F261**, **F035**, and **F054**). Of all eight fragments, six donate an H-bond via their protonated nitrogen atoms to Asp119, while **F278** faces the residue via its carboxamide oxygen in 2.7 Å distance. This would require Asp119 to be protonated (Fig. 4.3, i). In the crystal structure of **F051**, the fragment's pyridine ring addresses Asp119 in a similar fashion (Fig. 4.3, e). Using the webserver tool *chemicalize.org*,[25] the protonation state of the pyridine ring in solution is predicted as largely deprotonated at the applied soaking conditions. Thus, also in this case Asp119 would have to be protonated to establish the observed hydrogen bond. Even though neutron diffraction studies observe Asp119 to be deprotonated,[26,27] many studies have shown that the local environment can strongly influence the pK_a value of ionizable groups. Hence, the protonation state of Asp119 remains to be elucidated and might change in each individual case.[28,29]

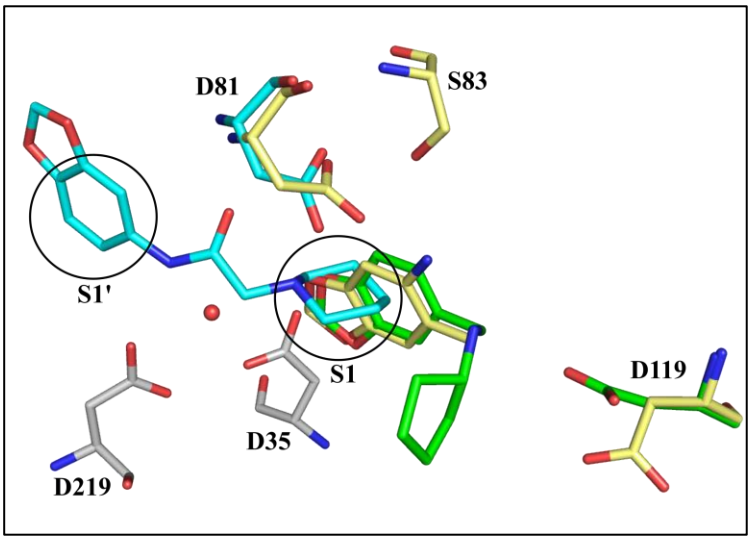


***Fig. 4.7**. Piperonyl moieties occupying different binding pockets. Yellow: **F323**, green: **F321**, cyan: **F041**.*

The eight remaining fragment molecules bind less specifically to the EP surface (18% of all remote binders, Fig. 4.5). Of these, four (**F031**, **F268**, **F131**, and **F261**) occupy the S1' pocket whereas the remaining four (**F114**, **F171**, **F056**, and **F227**) bind on the EP surface, outside of the substrate binding site. While two of the surface binders occupy also an additional position in the substrate binding site (**F114** and **F171)**, **F056** and **F227** appear only once in the respective crystal structures. **F227** is the only fragment among the wealth of fragment hits that forms a covalent attachment via Asp279 to EP near a disulfide bridge. Predictions of the pK_a value suggest **F056** and **F227** to be deprotonated at their acid functions, thus making plausible why binding sites remote from the rather negatively charged EP substrate-binding site appear comprehensible. Furthermore, **F261** attracts attention as it was considered a hit by the HCS and RDA prescreening methods from our previous study, and the subsequently performed ITC measurements further characterized this fragment as surprisingly efficient surface binder with a K_d value of 0.6 mM and an LE = 0.23 (kcal mol^{-1} $atom^{-1}$).[4,22]

The analyzed fragments and their binding poses are an excellent source of information for putative optimization strategies.

EP was and is still a relevant surrogate to investigate putative binding modes of potent peptidomimetic renin and ß-secretase inhibitors, long before the crystal structures of the latter enzymes became available. It is important, however, to consider deviations between the model and actual proteins. We thus compared the structures of EP and five additional therapeutically relevant pepsin-like aspartic proteases (renin, ß-secretase (BACE-1), secreted aspartic protease (SAP2), cathepsin D, and plasmepsin II).[30] Overall, the essential regions around the catalytic dyad are well conserved in all six structures. Furthermore, a tyrosine residue (Tyr79) in the flap region is commonly shared by all six proteases. One of the EP's hot spots Asp81 is only present in SAP2, whereas cathepsin D and plasmepsin II exhibit a serine at the same place. In renin, a serine is found at the preceding position 80. Moreover, Asp119 is shared by renin, cathepsin D, and plasmepsin II.

In a recent study on EP inhibitors, synthesized by the Gewald reaction, we could demonstrate, that small variations at the parent scaffold resulted in quite distinct inhibitor binding poses, not least owing to the large size of the EP binding site (Fig. 4.8, A).[31] Considering these ligands, expansion of their basic molecular scaffold toward the above-mentioned hot-spot residue Asp119

might improve binding. The phenyl or indolyl substituents, presently attached to their parent scaffold and accommodated in the S1 pocket, could potentially be substituted with a cyclopentanamine, as suggested by the binding pose of **F321** (Fig. 4.8, B). Striving for an additional hydrogen bond to be established with Asp119, a potential affinity enhancement can be expected. Phe291 however, hot spot in EP, does not occur in any of the other mentioned proteases at the same position. Instead, an aspartate residue is found at the same position in renin. Though, such a difference can be used to sustain purposefully designed ligands with a desired selectivity advantage for the target protein. If we analyze for instance, the binding pose of three early-stage renin inhibitors, which had been co-crystallized with EP and deposited in the protein databank, we see that they form weak hydrophobic CH-π contacts with the EP hot spot residue Phe291 using either a *tert*-butyl (PDB-codes: 2ER9, 2ER7), or *iso*-propyl group (PDB-code: 2ER0) (Fig. 4.9, A).[32–34] Furthermore, they place a proline residue into the S4 pocket. Accordingly, the P4 and P6 substituents remain spatially separated. For improved EP binding, our fragment study would suggest to merge the P4 and P6 substituents in the S4 and S6 pockets, simultaneously optimizing the interactions with the hot spot at Phe291.

Several fragments, e.g. **F034**, **F181**, **F273**, **F291**, **F311**, **F328**, and **F337** suggest such binding. These fragments extend across the S4 and S6 pockets and involve Phe291 in a favorable contact (Fig. 4.9, B).

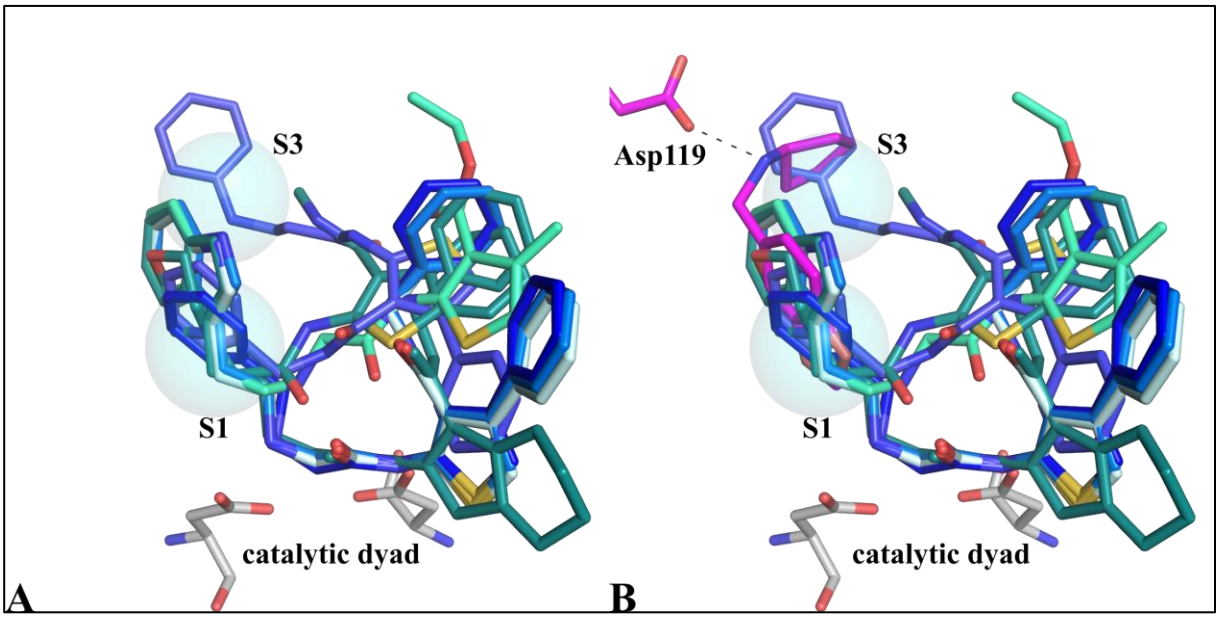


***Fig. 4.8**. **A**: Superposition of seven inhibitors, synthesized by the Gewald reaction, is shown by using graded blue colors for the ligands (PDB-codes: 3T7Q, 4L6B, 3PSY, 3WZ6, 3WZ7, 3T7X, and 3WZ8).[39] **B**: The inhibitors shown in A are superimposed with the crystal structure of our EP-**F321** complex (magenta).*

A similar concept is suggested to merge the occupants of the hydrophobic S1 and S3 pockets, which are filled by a P1-cyclohexylalanyl and a P3-phenylalanyl side chain in the crystal structures of many early-stage renin inhibitors.[33,35] With such a strategy, a breakthrough in the renin project was made at Ciba (now Novartis), which resulted in the marketed renin inhibitor Aliskiren.[36,37] Similarly, our detected fragments **F278**, **F321**, **F328**, and **F291** suggest such a promising bridging from the S1 into the S3 pockets (Fig. 4.9, B). The latter fragment **F291** would even be able to address the catalytic W501 via its pyridine-type nitrogen atom.

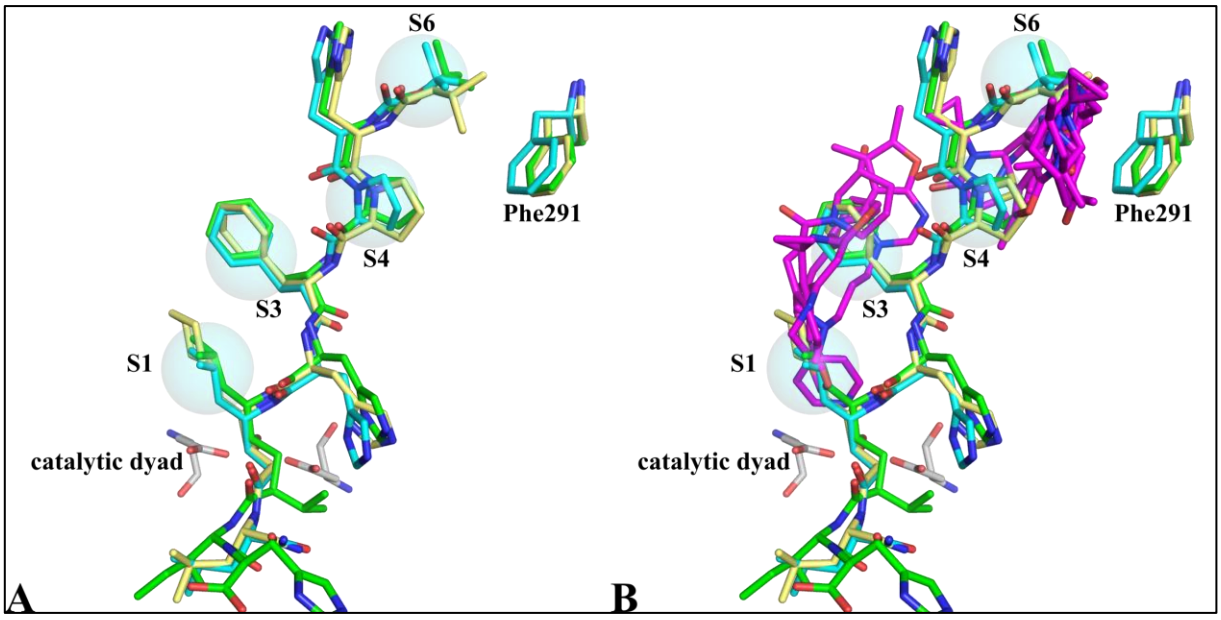


Fig. 4.9. A*: Superposition of three peptidomimetic renin inhibitors co-crystallized with EP. Binding pockets are indicated by cyan transparent spheres and heteroatoms are colored according to the usual color-coding. PDB-codes: 2ER9 (cyan), 2ER7 (green), and 2ER0 (yellow).*[32–34] ***B****: Superposition of the peptidomimetic inhibitors and the fragments found in the respective EP complexes.* ***F278****,* ***F291****,* ***F321****, and* ***F328*** *expand over the S1 and S3 pockets, whereas S4 and S6 are simultaneously occupied by* ***F034****,* ***F181****,* ***F273****,* ***F291****,* ***F311****,* ***F328****, and* ***F337****. For reasons of clarity, all fragment molecules are colored in magenta, with the usual color-coding for heteroatoms.*

Furthermore, in many peptidomimetic renin inhibitors an amide group is found between the S1' and S2' pockets. A similar pattern is detected for **F268** and **F131** at that position (Fig. 4.5, b and c). These fragments bridge between the two pockets addressing Gly37 and Gly80 with their amide groups. While Gly80 is only found in SAP2 and cathepsin D at the same position, the position of Gly37 is conserved throughout all discussed pepsin-like protein structures, suggesting the amide-bond motif as a very promising pattern for optimization.

These examples show that the plethora of fragments observed in this study can propose structural motifs actually found in larger ligands. They also suggest the fusion of occupants in adjacent

binding pockets emerged from the fragment's binding positions with respect to that of the natural substrate. The collected information is meant to serve as a template, based on which proposals for optimization strategies can be acquired.

4.6 Conclusions

The present study uncovered an impressive high number of binding modes by use of 29 fragments disclosing 45 distinct binding poses in regions remote from the catalytic center of the aspartic protease EP. The broad chemical diversity of the discovered fragment hits enables a meaningful mapping of the available chemical space. Fragments addressing the catalytic dyad have higher relevance as starting points for the *de novo* design of a lead structure. However, the remote binders constitute a valuable source of information on how to develop feasible lead-optimization strategies particularly with respect to selectivity. The large number of detected fragments indicates that only a small number of hot spots apart from the catalytic dyad is given. Most likely, a putative ligand scaffold has to address these spots successfully in order to achieve optimal binding with a limited number of iteration steps. However, these hot spots may differ in their positions in structures of other pepsin-like aspartic proteases. Thus, only global aspects can be gained from such analysis. With respect to physicochemical properties, we observe that larger fragments preferentially occupy remote parts of the substrate-binding site (Table 4.1 and 4.2). Furthermore, fragments with amino groups are preferred by aspartic proteases, considering that many members of this enzyme family operate in acidic conditions. Moreover, only two of the 20 primary aliphatic amines from our fragment library bind remotely, whereas this number increases gradually for secondary and tertiary amines, indicating once again that sterically more demanding fragments prefer remote binding sites.

Additionally, we detected fragments bridging between adjacent pockets, e.g., **F278**, **F321**, **F328**, and **F291** between the S1 and S3 pockets, or **F034**, **F181**, **F273**, **F291**, **F311**, **F328**, and **F337**, found to virtually crosslink S4 and S6. Such fragments can provide valuable ideas on how to modify and reduce the size of a peptidomimetic inhibitor during optimization campaigns. We thus believe that the large selection of remote binding fragments combined with an appropriate optimization strategy will significantly support the discovery of new attractive scaffolds but also the optimization of existing ones.

4.7 Experimental Section

Protein crystallization, soaking experiments, data collection, processing, and structure determination, graphical representation of the pockets in the binding site of EP as well as the computational predictions of fragment properties have already been described in chapter 3.7.

4.8 References

[1] Imming, P.; Sinning, C.; Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Discovery* **2006**, *5*, 821–834.

[2] Austin, D. J.; Crabtree, G. R.; Schreiber, S. L. Proximity versus allostery: the role of regulated protein dimerization in biology. *Chem. Biol.* **1994**, *1*, 131–136.

[3] Ena, J.; Amador, C.; Benito, C.; Pasquau, F. Pharmacological and clinical evidence of nevirapine immediate- and extended-release formulations. *HIV/AIDS* **2012**, *4*, 169–179.

[4] Schiebel, J.; Radeva, N.; Köster, H.; Metz, A.; Krotzky, T.; Kuhnert, M.; Diederich, W. E.; Heine, A.; Neumann, L.; Atmanene, C.; Roecklin, D.; Vivat-Hannah, V.; Renaud, J.-P.; Meinecke, R.; Schlinck, N.; Sitte, A.; Popp, F.; Zeeb, M.; Klebe, G. One question, multiple answers: biochemical and biophysical screening methods retrieve deviating fragment hit lists. *ChemMedChem* **2015**, *10*, 1511–1521.

[5] Neumann, L.; König, K. von; Ullmann, D. HTS reporter displacement assay for fragment screening and fragment evolution toward leads with optimized binding kinetics, binding selectivity, and thermodynamic signature. *Methods Enzymol.* **2011**, *493*, 299–320.

[6] Campos-Olivas, R. NMR screening and hit validation in fragment based drug discovery. *Curr. Top. Med. Chem.* **2011**, *11*, 43–67.

[7] Cala, O.; Krimm, I. Ligand-orientation based fragment selection in STD NMR screening. *J. Med. Chem.* **2015**, *58*, 8739–8742.

[8] Kranz, J. K.; Schalk-Hihi, C. Protein thermal shifts to identify low molecular weight fragments. *Methods Enzymol.* **2011**, *493*, 277–298.

[9] Linke, P.; Amaning, K.; Maschberger, M.; Vallee, F.; Steier, V.; Baaske, P.; Duhr, S.; Breitsprecher, D.; Rak, A. An automated microscale thermophoresis screening approach for fragment-based lead discovery. *J. Biomol. Screening* **2016**, *21*, 414–421.

[10] Cooper, J.; Quail, W.; Frazao, C.; Foundling, S. I.; Blundell, T. L.; Humblet, C.; Lunney, E. A.; Lowther, W. T.; Dunn, B. M. X-ray crystallographic analysis of inhibition of endothiapepsin by cyclohexyl renin inhibitors. *Biochemistry* **1992**, *31*, 8142–8150.

[11] Geschwindner, S.; Olsson, L.-L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; Beer, T. de; Folmer, R. H. A. Discovery of a novel warhead against beta-secretase through fragment-based lead generation. *J. Med. Chem.* **2007**, *50*, 5903–5911.

[12] Huschmann, F. U.; Linnik, J.; Sparta, K.; Ühlein, M.; Wang, X.; Metz, A.; Schiebel, J.; Heine, A.; Klebe, G.; Weiss, M. S.; Mueller, U. Structures of endothiapepsin–fragment complexes from crystallographic fragment screening using a novel, diverse and affordable 96-compound fragment library. *Acta Crystallogr. Sect. F* **2016**, *72*, 346–355.

[13] Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011**, *54*, 7784–7796.

[14] Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.

[15] Radeva, N.; Schiebel, J.; Wang, X.; Krimmer, S. G.; Fu, K.; Stieler, M.; Ehrmann, F. R.; Metz, A.; Rickmeyer, T.; Betz, M.; Winquist, J.; Park, A. Y.; Huschmann, F.; Weiss, M.; Müller, U.; Heine, A.; Klebe, G. Active site mapping of an aspartic protease by multiple fragment crystal structures: what can be learned for medicinal chemistry. *J. Med. Chem*. **2016**, *59*, 9743–9759.

[16] Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of secondary binding sites in proteins using fragment screening. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15910–15915.

[17] Bauman, J. D.; Harrison, Jerry Joe E K; Arnold, E. Rapid experimental SAD phasing and hot-spot identification with halogenated fragments. *IUCrJ* **2016**, *3*, 51–60.

[18] Bauman, J. D.; Patel, D.; Dharia, C.; Fromer, M. W.; Ahmed, S.; Frenkel, Y.; Vijayan, R. S. K.; Eck, J. T.; Ho, W. C.; Das, K.; Shatkin, A. J.; Arnold, E. Detecting allosteric sites of HIV-1 reverse transcriptase by X-ray crystallographic fragment screening. *J. Med. Chem.* **2013**, *56*, 2738–2746.

[19] Rashkin, M. J.; Waters, M. L. Unexpected substituent effects in offset pi-pi stacked interactions in water. *J. Am. Chem. Soc.* **2002**, *124*, 1860–1861.

[20] Bloom, J. W. G.; Wheeler, S. E. Taking the aromaticity out of aromatic interactions. *Angew. Chem.* **2011**, *123*, 7993–7995.

[21] Turk, J. A.; Smithrud, D. B. Synthesis and physical properties of protein core mimetics. *J. Org. Chem.* **2001**, *66*, 8328–8335.

[22] Schiebel, J.; Radeva, N.; Krimmer, S. G.; Wang, X.; Stieler, M.; Ehrmann, F. R.; Fu, K.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Heine, A.; Klebe, G. Six biophysical screening methods miss a large proportion of crystallographically discovered fragment hits: a case study. *ACS Chem. Biol.* **2016**, *11*, 1693–1701.

[23] Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem.* **2003**, *42*, 1210–1250.

[24] Burley, S. K.; Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **1986**, *203*, 139–143.

[25] Southan, C.; Stracz, A. Extracting and connecting chemical structures from text sources using chemicalize.org. *J. Cheminf.* **2013**, *5*, 20.

[26] Coates, L.; Erskine, P. T.; Wood, S. P.; Myles, D. A.; Cooper, J. B. A neutron Laue diffraction study of endothiapepsin: implications for the aspartic proteinase mechanism. *Biochemistry* **2001**, *40*, 13149–13157.

[27] Cooper, J. B.; Myles, D. A. A preliminary neutron Laue diffraction study of the aspartic proteinase endothiapepsin. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56*, 246–248.

[28] Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J. Biol. Chem.* **2009**, *284*, 13285–13289.

[29] Matthew, J. B.; Gurd, F. R.; Garcia-Moreno, B.; Flanagan, M. A.; March, K. L.; Shire, S. J. pH-dependent processes in proteins. *Crit. Rev. Biochem. Mol. Biol.* **1985**, *18*, 91–197.

[30] Rawlings, N. D.; Waller, M.; Barrett, A. J.; Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **2014**, *42* (Database issue), D503-509.

[31] Kuhnert, M.; Köster, H.; Bartholomäus, R.; Park, A. Y.; Shahim, A.; Heine, A.; Steuber, H.; Klebe, G.; Diederich, W. E. Tracing binding modes in hit-to-lead optimization: chameleon-like poses of aspartic protease inhibitors. *Angew. Chem.* **2015**, *54*, 2849–2853.

[32] Cooper, J. B.; Foundling, S. I.; Blundell, T. L.; Boger, J.; Jupp, R. A.; Kay, J. X-ray studies of aspartic proteinase-statine inhibitor complexes. *Biochemistry* **1989**, *28*, 8596–8603.

[33] Foundling, S. I.; Cooper, J.; Watson, F. E.; Cleasby, A.; Pearl, L. H.; Sibanda, B. L.; Hemmings, A.; Wood, S. P.; Blundell, T. L.; Valler, M. J. High resolution X-ray analyses of renin inhibitor-aspartic proteinase complexes. *Nature* **1987**, *327*, 349–352.

[34] Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L. X-ray analyses of aspartic proteinases. III Three-dimensional structure of endothiapepsin complexed with a transition-state isostere inhibitor of renin at 1.6 A resolution. *J. Mol. Biol.* **1990**, *216*, 1017–1029.

[35] Boger, J.; Lohr, N. S.; Ulm, E. H.; Poe, M.; Blaine, E. H.; Fanelli, G. M.; Lin, T.-Y.; Payne, L. S.; Schorn, T. W.; LaMont, B. I.; Vassil, T. C.; Stabilito, I. I.; Veber, D. F.; Rich, D. H.; Bopari, A. S. Novel renin inhibitors containing the amino acid statine. *Nature* **1983**, *303*, 81–84.

[36] Nussberger, J.; Delabays, A.; Gasparo, M. de; Cumin, F.; Waeber, B.; Brunner, H. R.; Menard, J. Hemodynamic and biochemical consequences of renin inhibition by infusion of CGP 38560A in normal volunteers. *Hypertension* **1989**, *13*, 948–953.

[37] Gradman, A. H.; Schmieder, R. E.; Lins, R. L.; Nussberger, J.; Chiang, Y.; Bedigian, M. P. Aliskiren, a novel orally effective renin inhibitor, provides dose-dependent antihypertensive efficacy and placebo-like tolerability in hypertensive patients. *Circulation* **2005**, *111*, 1012–1018.

[38] Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 251–264.

[39] Kuhnert, M.; Köster, H.; Bartholomäus, R.; Park, A. Y.; Shahim, A.; Heine, A.; Steuber, H.; Klebe, G.; Diederich, W. E. Tracing binding modes in hit-to-lead optimization: chameleon-like poses of aspartic protease inhibitors. *Angew. Chem. Int. Ed.* **2015**, *54*, 2849–2853.

5. Comparison of Cocktail versus Single Soaking Experiments: Pitfalls to be emphasized

5.1 Introductory remarks

The set up and data collection of cocktail experiments A and C and the single soaking experiment with **F333** at 50 mM were prepared by Johannes Schiebel. The set up and data collection of cocktails B and D were performed by Helene Köster.
The author of this thesis performed the structure refinement of all experiments provided by Johannes Schiebel and carried out the comparison of the data with the corresponding single fragment soaks described in chapters 3 and 4.

5.2 Introduction

X-ray crystallography is gaining in importance as primary screening method in FBDD since recent progress in academia and industry. [1–3] Synchrotron radiation, robotic crystal mounting and sample changing, advancement in detector technologies, and automated refinement pipelines have been introduced to increase the method's applicability as an efficient high-throughput technique.[1,4,5] In this context the method outperforms other techniques as it can handle higher fragment concentrations that other methods cannot. Furthermore, eventual ligand precipitation can be neglected as soon as the deposit is removed from the crystal surface immediately before freezing. Advantage of this procedure is that the fragment is exposed to a saturated solution and changes in concentration by absorption of the fragment into the crystal are easily compensated by increasing solvation.

A popular procedure is the design and application of cocktail experiments involving up to eight compounds in a mixture in order to accelerate the hit identification procedure.[6–9] The technique allows for the fast reduction of the number of non-binders while estimating bound fragments based on the electron density. However, the higher the number of compounds in a mixture, the higher the influence on parameters such as solubility and the thereof resulting reduced fragment occupancy, chemical reactivity, and crystal damage. Furthermore, compounds often prefer one and the same region for binding resulting in rather diffuse electron density, which is difficult to interpret or which requires subsequent analysis, decelerating the process of hit identification.

In such a case, single soaking experiments are subsequently carried out to determine the bound compound with absolute certainty. First of all, to assign the electron density unambiguously to one of the cocktail members, the structural diversity of all compounds included in a mixture must be well selected. Second, such a repetitive single soaking protocol is only applied if the initial cocktail soaking indicated any successful binding of a cocktail member.

The purpose of this chapter is to set out major differences between structures of singly soaked fragments and cocktails exposing these fragments as an example to the aspartic protease endothiapepsin (EP). The following examples will demonstrate how similar binding motifs compete for same binding sites and how the quality of electron densities is significantly influenced by the presence of two or more fragments in a mixture.

5.3 Results and Discussion

In the first experiment three fragments (**F63**, **F267**, and **F291**) with pyridine moieties were simultaneously soaked at a concentration of 90 mM each (cocktail A). The produced data of cocktail A are of high quality with statistics slightly better than the mean of the data sets collected for singly soaked fragments (Table 5.1).[10] In the single soaking experiments all three fragments addressed the catalytic water W501 and oriented toward the S1 pocket with their pyridine rings (Fig. 5.1, A–C). However, in the cocktail experiment varying binding sites for **F63** and for one molecule of **F291** were observed (Fig. 5.1, D–D_2). However, despite the simultaneous fragment soaking into EP, the difference electron density in the cocktail structure is well defined. This allowed the assignment of all fragments together with the doubly populated **F291**.

Table 5.1 *Data collection and refinement statistics cocktail* ***A*** *and single soaking experiments.*

Fragment number	**Cocktail A: F63, F267, F291**	**F63**	**F267**	**F291**
PDB code	**5MB0**	**4Y57**	**4Y5C**	**4Y45**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
Detector	PILATUS 6M	MX-225 CCD	PILATUS 6M	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.4, 73.1, 52.5	45.6, 73.3, 53.1	45.3, 73.1, 52.8	45.3, 73.0, 52.8
α, β, γ (°)	90.0, 109.6, 90.0	90.0, 109.9, 90.0	90.0, 109.4, 90.0	90.0, 109.6, 90.0
Resolution range (Å)	42.8-1.15 (1.22-1.15)	42.8-1.49 (1.58-1.49)	42.7-1.19 (1.26-1.19)	42.7-1.06 (1.12-1.06)
Wilson B factor ($Å^2$)	9.5	10.9	11.2	9.7
No. of unique reflections	113549 (17691)	53579 (8621)	103484 (16282)	144991 (22289)
Multiplicity	3.6 (3.2)	4.2 (4.2)	3.5 (3.3)	3.6 (3.3)
R_{sym} (%)	7.4 (46.6)	7.5 (49.4)	3.7 (49.4)	3.8 (46.5)
Completeness (%)	98.8 (95.2)	99.8 (99.6)	99.1 (96.8)	98.8 (94.1)
$\langle I/\sigma(I) \rangle$	9.5 (2.8)	13.5 (2.9)	17.9 (2.2)	16.7 (2.4)
Refinement				
Resolution range	42.8 - 1.15	41.2 - 1.49	39.5 - 1.19	42.7 - 1.06
No. of reflections used for refinement (work (free))	113540 (5677)	50900 (2679)	98309 (5175)	137741 (7250)
R_{cryst} (%)	15.2	11.7	11.5	12.0
R_{free} (%)	17.0	15.3	14.1	13.6
No. of refined residues	330	330	330	330
No. of fragment atoms	59	15	16	31
No. of other ligand atoms	30	20	22	18
No. of water molecules	387	326	277	288
RMSD, bond length (Å)	0.006	0.008	0.006	0.006
RMSD, bond angles (°)	0.9	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	94.2	93.5	93.5	92.8
Additionally allowed	5.8	6.5	6.5	7.2
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	11.2	11.8	13.2	11.5
Main chain	10.3	10.9	12.0	10.5
Side chain	12.1	12.6	14.4	12.4
Fragment atoms	13.4	13.2	34.7	16.1
Other ligand atoms [d]	21.6	24.6	26.5	16.2
Waters	28.7	27.2	27.6	24.9

[a] Statistic for the highest-resolution shell are shown in parenthesis; [b] Calculated by PROCHECK; [c] Calculated by MOLEMAN; [d] Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.

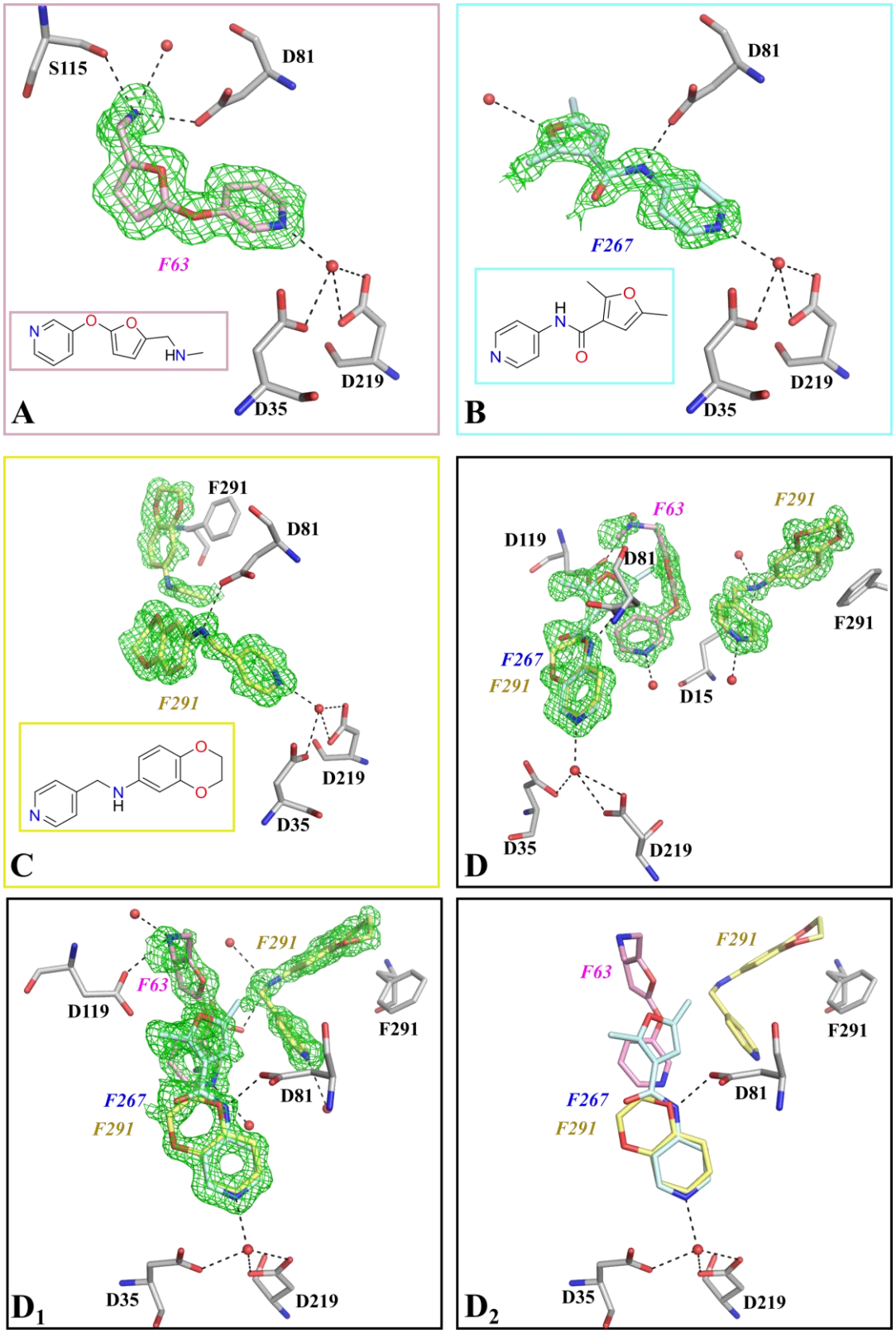


Fig. 5.1. *Crystal structures of the singly soaked fragments* ***F63***, ***F267***, *and* ***F291*** *(A-C) versus fragment binding from cocktail* ***A*** *(D, D_1, and D_2). For all heteroatoms, the standard color-coding has been used, while the carbon atoms have been colored based on the assigned fragment. The mF_o-DF_c meshes around the fragments are shown at 3σ.*

However, as not all pyridine moieties of the three fragments can simultaneously occupy the S1 pocket next to the catalytic W501 as observed in the single soaks, a competition must have occurred, leading to the spatial displacement of **F63** addressing now Asp119 with its secondary amine. **F267** was found to bind as expected in full agreement to the single soak. Also, one molecule of **F291** addressing Phe291 through π-π stacking is observed as in the single soak. However, the second molecule of **F291** could only be partly assigned to the electron density with its benzodioxanyl instead of its pyridine moiety in the S1 pocket.

In another cocktail experiment (cocktail B), a set of two fragments (**F224** and **F236**) was prepared at 50 mM (Table 5.2). Also here, based on the structural information obtained from the single soaking experiments, fragments with an aliphatic secondary amine as warhead within the set were chosen from the library. In the single soaking experiments these fragments occupy different binding pockets supposedly as no direct competition as given in the cocktail has been experienced (Fig. 5.2, A and B). Although both fragments use their secondary amino group for an H-bond interaction, **F224** places one of the bound copies under the flap region addressing Asp81 and the second molecule in van der Waals distance to the first one where its secondary amino group is virtually wrapped by an oligoethylene glycol picked up from the buffer. In contrast, **F236** uses its secondary amino group for a direct interaction with the catalytic dyad. An overlay of the initial fragment binding poses with the $mF_o–DF_c$ map at 3σ from cocktail B clearly demonstrates that the original binding modes are no longer preserved by both fragments (Fig. 5.2, C). Moreover, the resulting density allows only the placement of **F236** under the flap region, forming a 2.8 Å H-bond to Asp81 via its secondary amino group, where one of the molecules of **F224** in the single soaking experiment was found (Fig. 5.2, C). The residual electron density in the cocktail experiment at 3σ suggests that the binding mode found for **F236** (Fig. 5.2, B) in the single soak is no longer adopted, instead the fragment is moved next to Asp81. However, the fragment's mean B-value of 37.6 $Å^2$ at a structural resolution of 1.30 Å indicates the mobility of the fragment molecule and the probability of varying fragment atomic positions in different unit cells (Table 5.2). Remarkably, hardly any electron density next to the catalytic dyad or between Asp81 and Asp119 suggest the population of **F224** (Fig. 5.2, D).

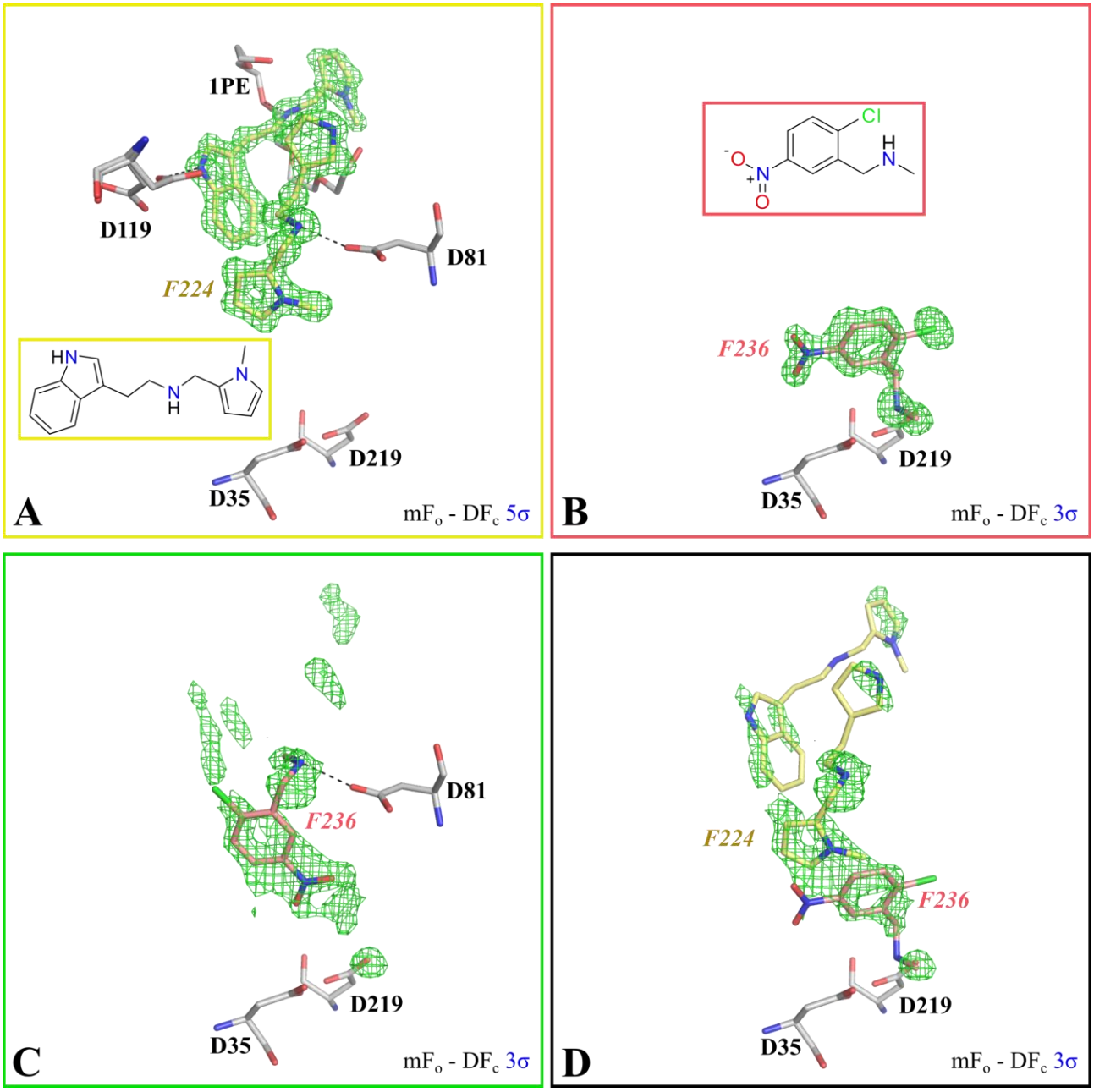


Fig. 5.2*. Crystal structures of the singly soaked fragments* ***F224*** *and* ***F236*** *(A and B) versus cocktail* ***B*** *(C), where only one fragment was modeled into the electron density. In D the original fragment binding modes are superimposed with the electron density, resulted from the cocktail experiment. For all heteroatoms, the standard color-coding has been used, while the carbon atoms have been colored based on the assigned fragment.*

Table 5.2 *Data collection and refinement statistics cocktail* ***B*** *and single soaking experiments.*

Fragment number	**Cocktail B: F224 and F236**	**F224**	**F236**
PDB code	**5MB7**	**4YD3**	**4YD5**
Data collection and processing			
Wavelength (Å)	0.91841	0.91841	0.91841
Beamline	BESSY BL14.2	BESSY BL14.1	BESSY BL14.1
Detector	MX-225 CCD	PILATUS 6M	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]			
a, b, c (Å)	45.4, 72.9, 52.7	45.3, 73.2, 52.7	45.3, 73.0, 52.7
α, β, γ (°)	90.0, 109.5, 90.0	90.0, 109.7, 90.0	90.0, 109.3, 90.0
Resolution range (Å)	30.0-1.30 (1.32-1.30)	42.7-1.25 (1.32-1.25)	42.8-1.21 (1.28-1.21)
Wilson B factor ($Å^2$)	10.9	9.8	9.8
No. of unique reflections	78870 (3975)	86779 (13216)	96832 (14290)
Multiplicity	3.2 (3.1)	3.4 (3.3)	3.6 (2.8)
R_{sym} (%)	5.2 (52.0)	3.9 (33.1)	5.2 (22.4)
Completeness (%)	99.9 (99.3)	96.5 (91.0)	98.1 (89.8)
$\langle I/\sigma(I) \rangle$	22.3 (2.0)	18.9 (3.3)	13.7 (3.6)
Refinement			
Resolution range	18.9 - 1.30	42.7 - 1.25	42.8 - 1.21
No. of reflections used for refinement (work (free))	78840 (3850)	82440 (4339)	91990 (4842)
R_{cryst} (%)	13.5	11.1	11.9
R_{free} (%)	15.8	13.2	13.6
No. of refined residues	330	330	330
No. of fragment atoms	13	34	13
No. of other ligand atoms	6	59	34
No. of water molecules	337	293	226
RMSD, bond length (Å)	0.006	0.006	0.006
RMSD, bond angles (°)	0.9	1.2	1.2
Ramachandran plot (%) [b]			
Most favoured	93.5	94.2	94.2
Additionally allowed	6.5	5.8	5.8
Generously allowed	0	0	0
Disallowed	0	0	0
Mean B-values ($Å^2$) [c]			
All protein atoms	12.5	11.4	11.6
Main chain	11.5	10.4	10.6
Side chain	13.4	12.3	12.5
Fragment atoms	37.6	15.6	18.6
Other ligand atoms [d]	17.5	23.6	23.6
Waters	27.2	27.0	25.5

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

The examples so far clearly demonstrate that the presence of two or more competing fragment binders can strongly influence the binding poses observed for separately studied fragments in single soaking experiments. Moreover, the examples highlight the importance of structural diversity when compiling cocktail experiments.

Otherwise same or very similar structural motifs would compete for one and the same binding site, often leading to diffuse and inexplicable electron density.

Thus, an additional cocktail experiment (cocktail C) was performed using a set of two fragments at 90 mM each. This time their structural diversity was regarded (Table 5.3). Thereby, the hydrazide **F103** and the thiazolamine **F171** were simultaneously soaked into EP. In the previously performed single soaking experiments **F171** occurred twice in the crystal as one molecule was found on the protein surface while the second one addressed Asp35 and the catalytic W501 with its exocyclic amino group (Fig. 5.3, A). In contrast, **F103** displaced the catalytic W501 while directly interacting with both of the catalytic aspartates (Fig. 5.3, B).

In cocktail C only the remotely bound molecule of **F171** could be clearly assigned to the electron density (Fig. 5.3, C).

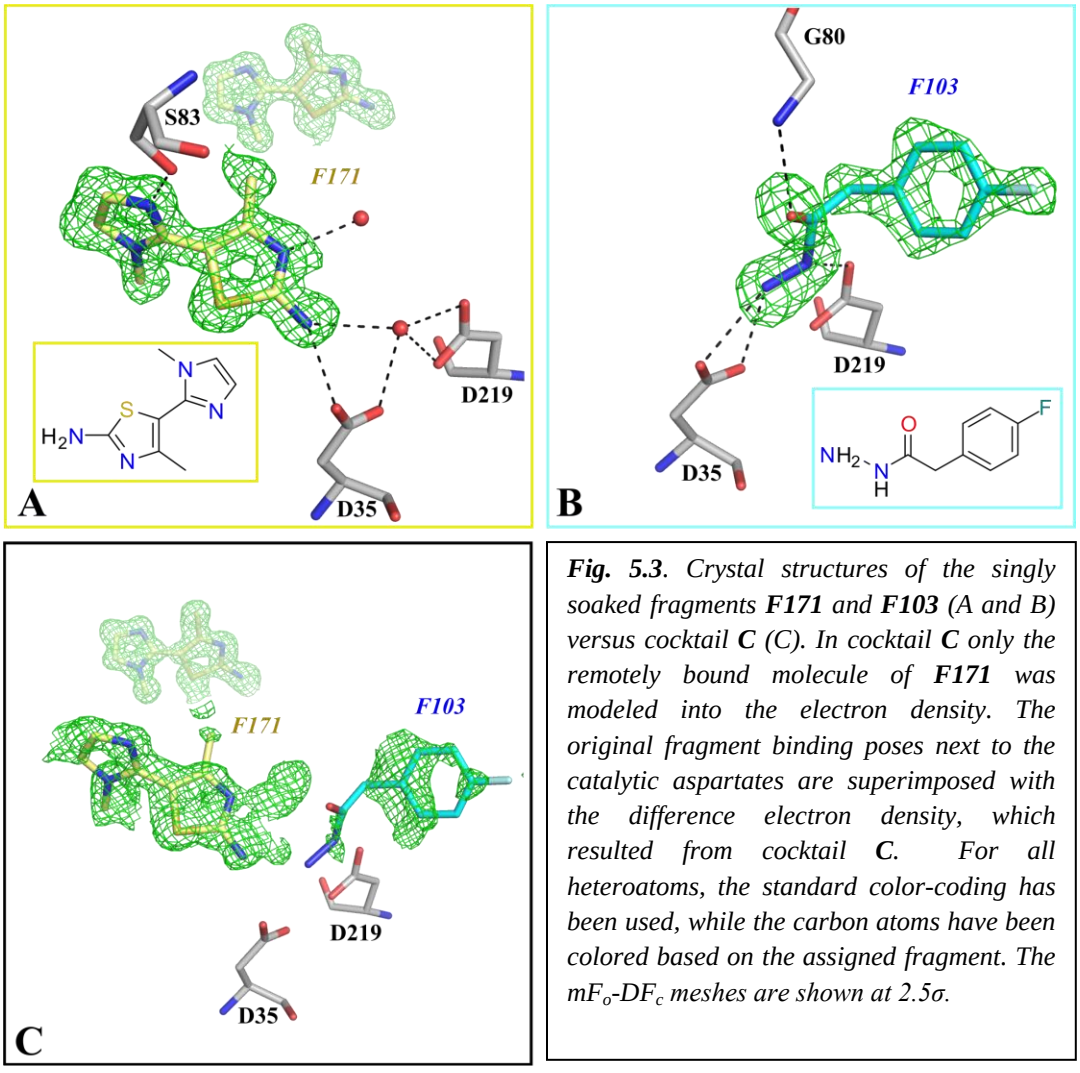


***Fig. 5.3**. Crystal structures of the singly soaked fragments **F171** and **F103** (A and B) versus cocktail **C** (C). In cocktail **C** only the remotely bound molecule of **F171** was modeled into the electron density. The original fragment binding poses next to the catalytic aspartates are superimposed with the difference electron density, which resulted from cocktail **C**. For all heteroatoms, the standard color-coding has been used, while the carbon atoms have been colored based on the assigned fragment. The mF_o-DF_c meshes are shown at 2.5σ.*

Table 5.3 *Data collection and refinement statistics cocktail* ***C*** *and single soaking experiments.*

Fragment number	**Cocktail C: F171 and F103**	**F171**	**F103**
PDB code	**5MB5**	**4Y3X**	**4Y3M**
Data collection and processing			
Wavelength (Å)	0.91841	0.91841	0.91841
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.2
Detector	PILATUS 6M	PILATUS 6M	MX-225 CCD
Space group	$P2_1$	$P2_1$	$P2_12_12_1$
Unit-cell parameters [a]			
a, b, c (Å)	45.4, 72.9, 52.6	45.3, 73.0, 52.7	45.2, 72.5, 104.1
α, β, γ (°)	90.0, 109.7, 90.0	90.0, 109.7, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	42.7-0.98 (1.04-0.98)	42.7-1.25 (1.33-1.25)	45.2-1.55 (1.64-1.55)
Wilson B factor ($Å^2$)	8.4	8.8	11.1
No. of unique reflections	179855 (27996)	87058 (13331)	50445 (8029)
Multiplicity	3.8 (3.5)	3.8 (3.6)	8.0 (8.1)
R_{sym} (%)	6.6 (45.0)	5.9 (40.0)	6.7 (42.0)
Completeness (%)	97.4 (94.0)	97.5 (92.7)	99.9 (99.7)
$\langle I/\sigma(I) \rangle$	10.2 (2.2)	13.9 (3.0)	23.1 (5.1)
Refinement			
Resolution range	36.5 – 0.98	42.7 - 1.25	42.3 - 1.55
No. of reflections used for refinement (work (free))	179793 (8989)	82705 (4353)	47922 (2523)
R_{cryst} (%)	14.7	12.0	14.5
R_{free} (%)	16.2	14.2	16.6
No. of refined residues	330	330	330
No. of fragment atoms	13	26	12
No. of other ligand atoms	30	32	43
No. of water molecules	408	296	294
RMSD, bond length (Å)	0.005	0.015	0.008
RMSD, bond angles (°)	0.9	1.6	1.2
Ramachandran plot (%) [b]			
Most favoured	93.5	93.5	93.1
Additionally allowed	6.5	6.5	6.9
Generously allowed	0	0	0
Disallowed	0	0	0
Mean B-values ($Å^2$) [c]			
All protein atoms	9.6	9.9	12.4
Main chain	8.8	9.1	11.4
Side chain	10.3	10.7	13.4
Fragment atoms	10.1	13.0	20.5
Other ligand atoms [d]	19.8	20.7	35.6
Waters	27.2	24.4	25.1

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Furthermore, **F103** can hardly be assigned to the electron density with the required confidence (Fig. 5.3, C). Although the resolution of the cocktail structure C with 0.98 Å clearly exceeds that of the single soaking experiments (1.25 and 1.55 Å) the electron density in the active site of cocktail C is significantly more diffuse and disconnected compared to the density of the single soaks.

We finally decided to remove the active site copy of **F171**, initially modeled into the density after subsequent refinement cycles, as a high level of negative and positive difference electron density remained after reducing and refining the fragment's occupancy.

In this experiment with an electron density from an optimally diffracting crystal, the refinement indicates a bound fragment in the active site but likely with strongly reduced occupancy. **F103** would have definitely failed to be discovered as hit without performing the study with individually soaked crystals, while the assignment of **F171** to the electron density strongly depends on how rigorously the individual crystallographer sets the criteria to interpret the residual electron density. Nevertheless, the single soaking experiment for **F171** would have provided an unambiguous answer. The cocktail experiment would have again greatly decelerated the hit identification process.

A similar experience has been made in our group in a previous study on EP, in which the in-house fragment library was first screened via a biochemical assay resulting in 55 hits, which were subsequently analyzed via X-ray crystallography using cocktails of two fragments at 50mM each.[9] Thereby, only structurally diverse fragments have been combined. In one of the cocktails, the benzothiopyranone **F308** and the pyrimidinone **F333** with an attached isopropanylthiolether were simultaneously soaked into *apo* EP crystals (cocktail D, Fig. 5.4, A). Although both fragments were identified as hits in the initially performed biochemical assay none of them could be found in the electron density of the studied cocktail D. Only a very fragmented density could be seen next to Phe291, not suited to unambiguously assign a fragment binder (Fig. 5.4, B). However, the subsequently performed single soaking experiments with **F333** at 50 mM (Fig. 5.4, C) and 90 mM (Fig. 5.4, D) clearly confirmed the fragment's binding next to Phe291.[11] In the single soaking experiment at a lower concentration, **F333** could be clearly assigned to the electron density, whereas the electron density of cocktail D next to Phe291 is hardly significant (Fig. 5.4, B). Also in this case, when the two fragments would have been selected randomly for a cocktail experiment and no complementary information regarding their potency would have been available both would have remained undetected as hits.

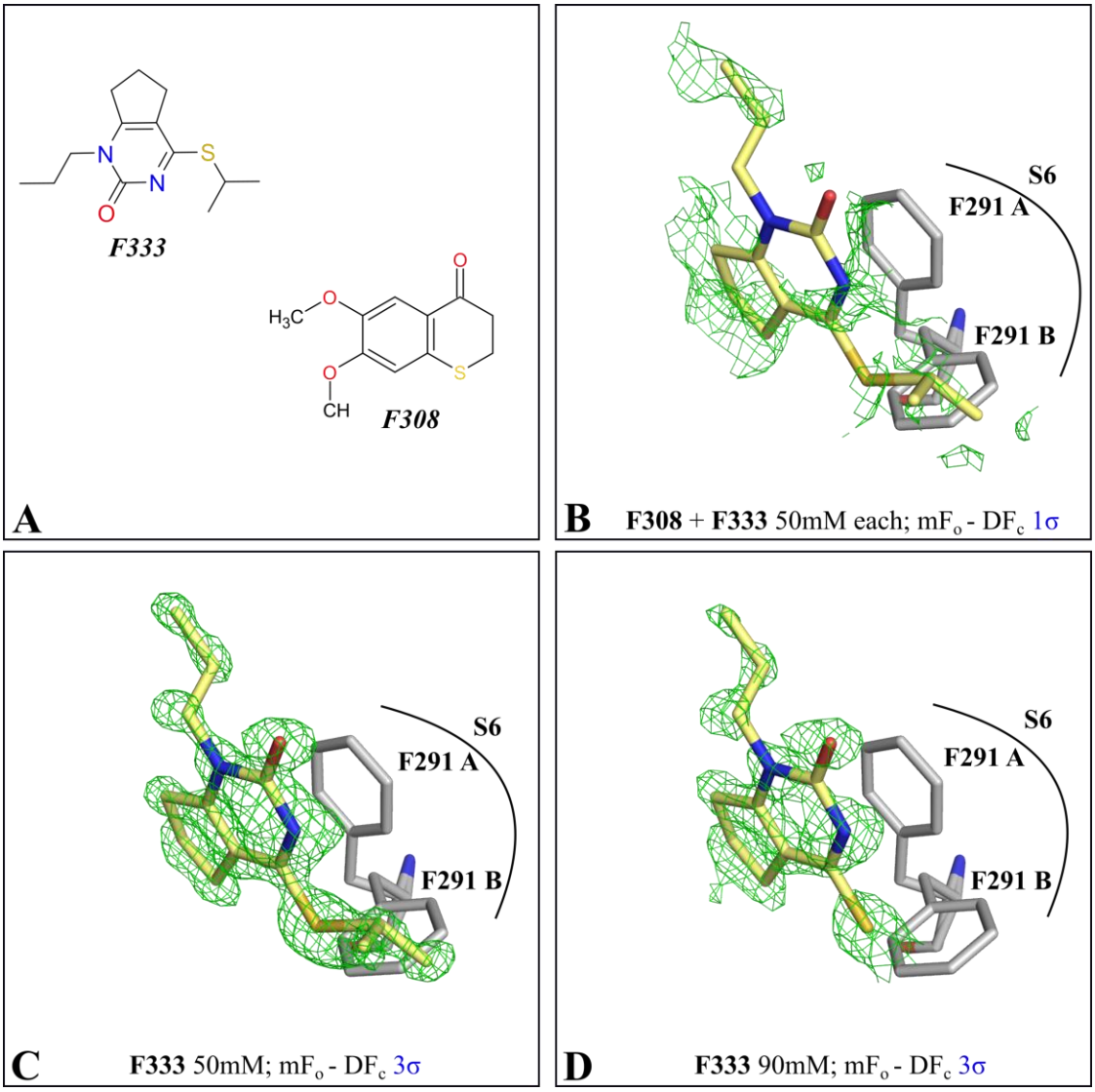


***Fig. 5.4**. Crystal structures of the singly bound fragment **F333** at concentrations of 50 mM (C) and 90 mM (D) versus cocktail **D** (B). The two dimensional structures of fragments **F333** and **F308**, used for the cocktail experiment, are presented in the upper left hand side of the figure (A). For all heteroatoms, the standard color-coding has been used, while the carbon atoms have been colored based on the assigned fragment.*

Table 5.4 *Data collection and refinement statistics cocktail* ***D*** *and single soaking experiments.*

Fragment number	**Cocktail D: F308 and F333**	**F333_50mM**	**F333_90mM**
PDB code	**5MB6**	**5MB3**	**5DR3**
Data collection and processing			
Wavelength (Å)	0.91841	0.91841	0.89500
Beamline	BESSY BL14.2	BESSY BL14.1	BESSY BL14.3
Detector	MX-225 CCD	PILATUS 6M	Mar CCD 165
Space group	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]			
a, b, c (Å)	45.4, 73.5, 53.2	45.4, 73.0, 52.7	45.3, 73.0, 52.8
α, β, γ (°)	90.0, 110.5, 90.0	90.0, 109.5, 90.0	90.0, 109.5, 90.0
Resolution range (Å)	30.0-1.20 (1.22-1.20)	42.8-1.05 (1.11-1.05)	42.7-1.24 (1.31-1.24)
Wilson B factor ($Å^2$)	8.3	9.1	9.6
No. of unique reflections	95743 (3095)	146432 (22858)	91107 (14458)
Multiplicity	2.8 (1.8)	3.8 (3.7)	4.2 (4.1)
R_{sym} (%)	5.4 (23.8)	5.0 (47.8)	4.4 (43.3)
Completeness (%)	94.2 (61.2)	97.0 (93.9)	99.2 (97.8)
$<I/\sigma(I)>$	21.7 (3.8)	13.0 (2.4)	20.3 (3.3)
Refinement			
Resolution range	27.9 – 1.20	36.5 - 1.05	20.9 - 1.24
No. of reflections used for refinement (work (free))	95693 (4629)	146410 (7319)	86551 (4556)
R_{cryst} (%)	13.3	13.3	11.2
R_{free} (%)	14.9	15.1	13.3
No. of refined residues	330	330	330
No. of fragment atoms	–	17	14
No. of other ligand atoms	11	28	28
No. of water molecules	464	407	341
RMSD, bond length (Å)	0.006	0.005	0.006
RMSD, bond angles (°)	0.9	0.9	1.2
Ramachandran plot (%) [b]			
Most favoured	94.2	93.5	93.5
Additionally allowed	5.8	6.5	6.5
Generously allowed	0	0	0
Disallowed	0	0	0
Mean B-values ($Å^2$) [c]			
All protein atoms	9.6	10.9	12.0
Main chain	8.7	10.1	11.0
Side chain	10.4	11.7	12.9
Fragment atoms	–	16.0	17.5
Other ligand atoms [d]	12.6	26.7	29.1
Waters	26.9	29.3	26.6

[a] Statistic for the highest-resolution shell are shown in parenthesis; [b] Calculated by PROCHECK; [c] Calculated by MOLEMAN; [d] Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.

5.4 Conclusions

In drug discovery, the information about the three dimensional structure of a fragment bound to its target is mandatory as it is prerequisite for any rational design in a hit-to-lead optimization campaign. X-ray crystallography can be used as a primary screening technique due to recent improvements in this field and the broad accessibility of synchrotron radiation. However, we believe that the application of cocktail experiments to maximize the method's throughput can easily lead to drop outs owing to false negative hit rates. Moreover, diffuse electron density results from reduced fragment occupancy provoked by lower fragment concentrations in the crystal and the mutual competition of fragments for same binding sites. Such densities often require additional analysis and, as a result, the hit identification process is decelerated.

Many of the fragments, which failed to be identified as hits in the cocktail experiments (**F224**, **F103**, one molecule of **F171**, and **F333**) appeared as clearly bound to EP using individual soaks. These results indicate that the individual fragment concentration in a cocktail is significantly lower than expected. Moreover, same warheads competing for the same binding sites result in either remote fragment binding (cocktail A, **F63**) or complete fragment displacement (cocktail B, **F224**). Furthermore, the kinetics of fragment penetration into crystals can take an important impact on the fragment population across a crystal. This population is not necessarily uniform across the entire specimen. In addition, the simultaneous presence of more than one fragment in a binding site might create novel contacts among the fragments influencing in a kind of synergistic effect the potency and the adopted binding pose of fragments. It is difficult to estimate how such observations can be exploited in fragment-to-lead optimization.

Since the complexity of fragment screening and binding pose characterization is clearly augmented by cocktail experiments, we strongly endorse the application of single soaking experiments instead of cocktails when consulting X-ray crystallography. To our opinion this is the only way to significantly reduce the false negative rates while optimizing the yield of well-defined electron densities to elucidate the three-dimensional structure of a fragment, bound to its target.

5.5 Experimental Section

The protein crystallization procedure has been described in chapter 3.7.1. Also, the implementation of cocktail experiments B (**F224** and **F236** at 50 mM each) and D (**F308** and **F333** at 50 mM each) has been previously descrived.[9] In short, the beforehand prepared crystals were soaked for 1-2 days in the reservoir buffer containing 0.1 M NH_4Ac, 0.1 M NaAc, and 26% (w/v) PEG 4000, to which 25% GOL and the two corresponding fragments at 50 mM each were added. After 48 h the crystals were frozen in liquid nitrogen. The diffraction data were collected at beamline 14.2 of the BESSY synchrotron in Berlin at 100 K and processed with *HKL2000*.[12] The structures were determined as described previously.[9]

The single soaking experiment with **F333** at a concentration of 50 mM was performed in 70 mM NH_4Ac, 70 mM NaAc, 16-20% (w/v) PEG 4000, and 23% GOL. The crystal was soaked for 48 h before freezing in liquid nitrogen.

Cocktail experiment A included three fragments (**F63**, **F267**, and **F291**) at 90 mM each. These were soaked for 24 h in a mixture of 65 mM NH_4Ac, 65 mM NaAc, 14-18% (w/v) PEG 4000, and 8% GOL. Cocktail experiment C (**F171** and **F103** at 90 mM each) was prepared analogous to cocktail A with the only difference that 17% instead of 8% GOL were used.

Diffraction data for cocktails A and C and for the single soaking experiment with **F333** at 50 mM were collected at BESSY beamline 14.1 and processed with XDS.[13] For the initial refinement of these datasets an automated refinement pipeline was applied.[1] For the molecular replacement step of the pipeline, the coordinates of the *apo* EP structure (PDB-code: 4Y5L) were used. In addition, the intermittent model building was performed with *Coot* as fragments and solvent molecules were modeled into a difference electron density of at least 2.5 σ.[14]

As very high structural resolutions are given, the structures were refined anisotropically, leading to significant improvement of the R_{free} values. If the level of difference electron density at appropriate ligand occupancy was significantly reduced after subsequent refinement cycles with *Phenix*, the ligand was retained.[15] In the final refinement cycle hydrogen atoms were added to the models and ADP weight were optimized. The mF_o-DF_c difference electron densities shown in the structural figures were generated with *Phenix*. All structural figures were made with *Pymol*.[16]

5.6 References

[1] Schiebel, J.; Krimmer, S. G.; Rower, K.; Knorlein, A.; Wang, X.; Park, A. Y.; Stieler, M.; Ehrmann, F. R.; Fu, K.; Radeva, N.; Krug, M.; Huschmann, F. U.; Glockner, S.; Weiss, M. S.; Mueller, U.; Klebe, G.; Heine, A. High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* **2016**, *24*, 1398–1409.

[2] Mooij, W. T. M.; Hartshorn, M. J.; Tickle, I. J.; Sharff, A. J.; Verdonk, M. L.; Jhoti, H. Automated protein-ligand crystallography for structure-based drug design. *ChemMedChem* **2006**, *1*, 827–838.

[3] Cohen, A. E.; Ellis, P. J.; Miller, M. D.; Deacon, A. M.; Phizackerley, R. P. An automated system to mount cryo-cooled protein crystals on a synchrotron beam line, using compact sample cassettes and a small-scale robot. *J. Appl. Crystallogr.* **2002**, *35*, 720**–**726.

[4] Henrich, B.; Bergamaschi, A.; Broennimann, C.; Dinapoli, R.; Eikenberry, E. F.; Johnson, I.; Kobas, M.; Kraft, P.; Mozzanica, A.; Schmitt, B. PILATUS: A single photon counting pixel detector for X-ray applications. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2009**, *607*, 247–249.

[5] Badger, J. Crystallographic fragment screening. *Methods Mol. Biol.* **2012**, *841*, 161–177.

[6] Nienaber, V. L.; Richardson, P. L.; Klighofer, V.; Bouska, J. J.; Giranda, V. L.; Greer, J. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* **2000**, *18*, 1105–1108.

[7] Sanders, W. J.; Nienaber, V. L.; Lerner, C. G.; McCall, J. O.; Merrick, S. M.; Swanson, S. J.; Harlan, J. E.; Stoll, V. S.; Stamper, G. F.; Betz, S. F.; Condroski, K. R.; Meadows, R. P.; Severin, J. M.; Walter, K. A.; Magdalinos, P.; Jakob, C. G.; Wagner, R.; Beutel, B. A. Discovery of potent inhibitors of dihydroneopterin aldolase using CrystaLEAD high-throughput X-ray crystallographic screening and structure-directed lead optimization. *J. Med. Chem.* **2004**, *47*, 1709–1718.

[8] Bauman, J. D.; Patel, D.; Dharia, C.; Fromer, M. W.; Ahmed, S.; Frenkel, Y.; Vijayan, R. S. K.; Eck, J. T.; Ho, W. C.; Das, K.; Shatkin, A. J.; Arnold, E. Detecting allosteric sites of HIV-1 reverse transcriptase by X-ray crystallographic fragment screening. *J. Med. Chem.* **2013**, *56*, 2738–2746.

[9] Koster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011**, *54*, 7784–7796.

[10] Radeva, N.; Schiebel, J.; Wang, X.; Krimmer, S. G.; Fu, K.; Stieler, M.; Ehrmann, F. R.; Metz, A.; Rickmeyer, T.; Betz, M.; Winquist, J.; Park, A. Y.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Heine, A.; Klebe, G. Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads To Address a Catalytic Dyad. *J. Med. Chem.* **2016**, *59*, 9743–9759.

[11] Radeva, N.; Krimmer, S. G.; Stieler, M.; Fu, K.; Wang, X.; Ehrmann, F. R.; Metz, A.; Huschmann, F. U.; Weiss, M. S.; Mueller, U.; Schiebel, J.; Heine, A.; Klebe, G. Experimental Active-Site Mapping by Fragments: Hot Spots Remote from the Catalytic Center of Endothiapepsin. *J. Med. Chem.* **2016**, *59*, 7561–7575.

[12] Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **1997**, 307–326.

[13] Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125–132.

[14] Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.

[15] Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213–221.

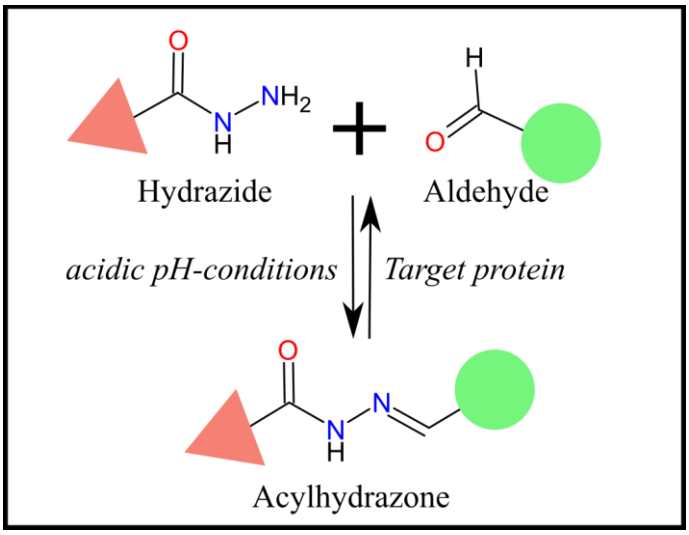
[16] DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14–20.

6. Structure-based Drug Design and Optimization of Inhibitors for Endothiapepsin by Exploiting Dynamic Combinatorial Chemistry

6.1 Introductory Remarks

The following work was done in cooperation with the group of Prof. Anna Hirsch from the University of Groningen, Netherlands, and coworkers from the National Hellenic Research Foundation in Athens, Greece, and the Stratingh Institute for Chemistry in Groningen, Netherlands. The work resulted in three shared publications: two were published in *Angewandte Chemie* in 2014 (DOI: 10.1002/anie.201309682) and 2016 (DOI: 10.1002/anie.201603074), and the third one was published in *Internatinal Journal of Molecular Science* in 2015 (DOI: 10.3390/ijms160819184).

The author of this thesis contributed the X-ray analysis of EP with bound ligands and the interpretation of the binding modes of those ligands to the publications. In this chapter the main content of the papers will be summarized whereby the main focus will be set on the structural information.



6.2 Introduction

Dynamic combinatorial chemistry (DCC) is a technique used for the molecular amplification of the members of a compound library assisted by the presence of a target protein. [1–3] The principle is based on the reversible and continuous interchange of the library members (building blocks) under thermodynamic control and in case of the studied system under acidic pH conditions. Such a dynamic combinatorial library (DCL) is therefore able to respond to molecular recognition by the target protein, which then can induce a shift of equilibrium, favoring the predominant formation of the selected species.

In the first project, we applied the technique to identify the best binder(s) from a DCL composed of five aldehydes and five hydrazides, reacting under a simple condensation reaction to acylhydrazones (Fig. 6.1).[4] First, different acylhydrazones were designed using the software MOLOC[5] and docked via FlexX[6], whereby two alternative binding poses were modeled, with and without binding through the catalytic water molecule. An additional primary amino group was introduced to serve as a putative anchor to bind to the target protein EP (see Chapter 3, warheads). Based on the docking analysis, a series of acylhydrazones were selected, which retrosynthetically led to the five hydrazides and five aldehydes shown in Fig. 6.1. Hydrazides H1 to H4 were subsequently synthetized from their enantiomerically pure methyl esters, while H5 and all five aldehydes were commercially available.

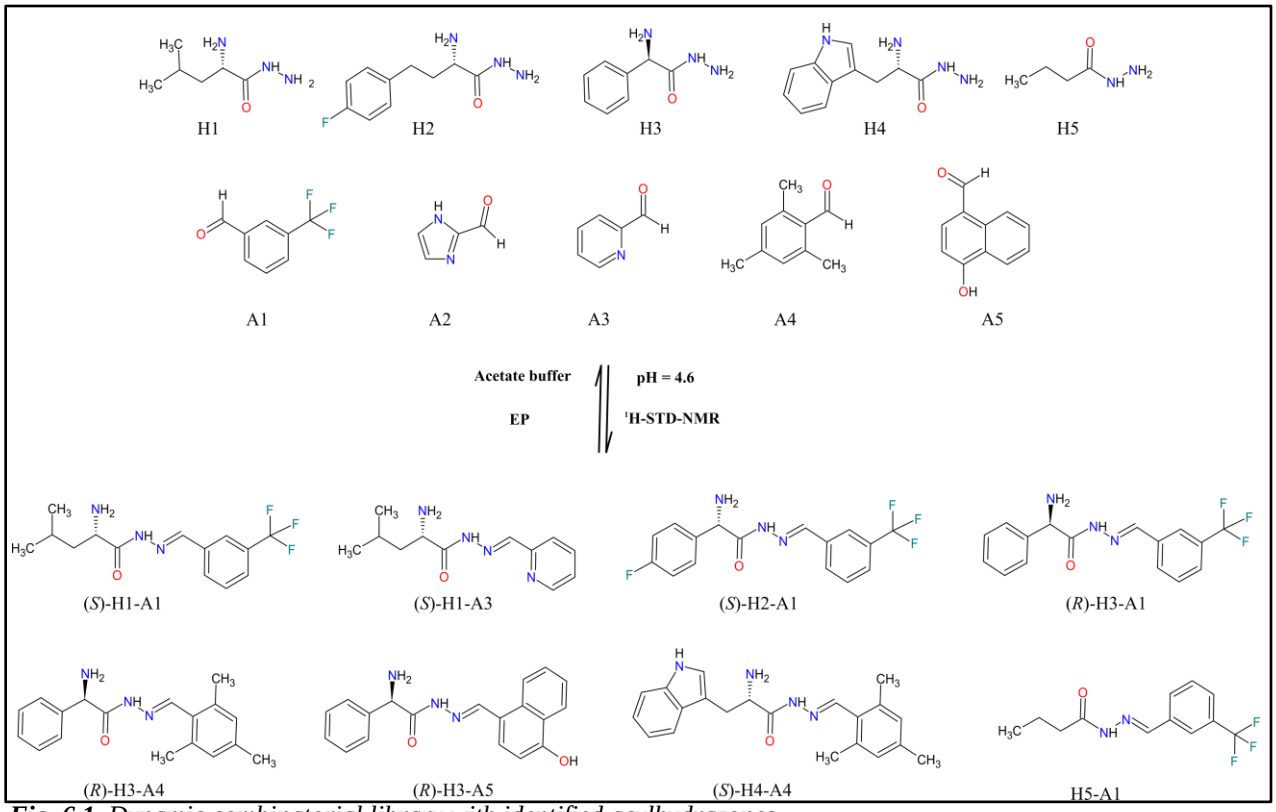


***Fig. 6.1**. Dynamic combinatorial library with identified acylhydrazones.*

For the condensation reaction, five sub-libraries were built, consisting of five hydrazides and one aldehyde each. Those sub-libraries were pre-equilibrated and upon addition of EP, eight protein-bound acylhydrazones were identified via ^{1}H-STD-NMR (Fig. 6.1).[7] The latter identified acylhydrazones were subsequently synthetized individually from their parent building blocks and their biological activity was tested in a fluorescence-based enzyme-kinetic assay (Table 6.1).

***Table 6.1**. Affinity data of the detected acylhydrazones.*

Inhibitors	IC_{50} (µM)	K_i (µM)	ΔG^0 (kJ·mol^{-1})[a]	LE (kJ·mol^{-1}·atom^{-1})
(*S*)-H4-A4	12.8 ± 0.4	6 ± 0.2	–30	0.27
(*R*)-H3-A5	14.5 ± 0.5	7 ± 0.2	–30	0.29
(*S*)-H1-A1	150 ± 17	71 ± 8	–24	0.27
(*S*)-H2-A1	177 ± 13	83 ± 6	–23	0.23
(*R*)-H3-A4	206 ± 19	97 ± 9	–23	0.25
(*R*)-H3-A1	352 ± 13	166 ± 6	–22	0.22
(*S*)-H1-A3	365 ± 95	172 ± 45	–22	0.30
H5-A1	insoluble	–	–	–

[a] *The Gibbs free energy of binding was derived from the experimentally determined IC_{50} values.*

To confirm the binding modes of the initially predicted binding poses, the acylhydrazones with the highest affinity, (*S*)-H4-A4 and (*R*)-H3-A5, were subsequently soaked into *apo* EP crystals and their binding modes were determined (e.g. 6.3 Results and Discussion).

We applied the method in a second round of design along with fragment linking strategy to improve the affinity and ligand efficiency of the acylhydrazones presented in the first project.[8] For this purpose the mesityl moiety of (*S*)-H4-A4 was virtually linked with the naphthyl moiety of (*R*)-H3-A5, whereby series of bis-acylhydrazones were designed. Retrosynthesis of the latter led to a bis-aldehyde (Fig. 6.2, bis-A1) and nine hydrazides (Fig. 6.2, 2-10). The bis-aldehyde and hydrazides 8 to 10 were purchased, while the remaining hydrazides (2-7) were synthetized from their corresponding methyl esters. The analysis was facilitated while combining the building blocks into two sub-libraries, each consisting of the bis-aldehyde and four or five hydrazides. In addition, a mixture of the bis-aldehyde with all nine hydrazides was made. All built bis-acylhydrazones were identified via LC-MS. However, upon EP addition to the pre-equilibrated sub-libraries, the amplification of six bis-acylhydrazones shown in Fig. 6.2 was observed. In addition, the bis-acylhydrazones bis-1 and bis-4 (Fig. 6.2) were synthetized from their corresponding building blocks and their biochemical activity was determined in a fluorescence-based assay (Table 6.2).

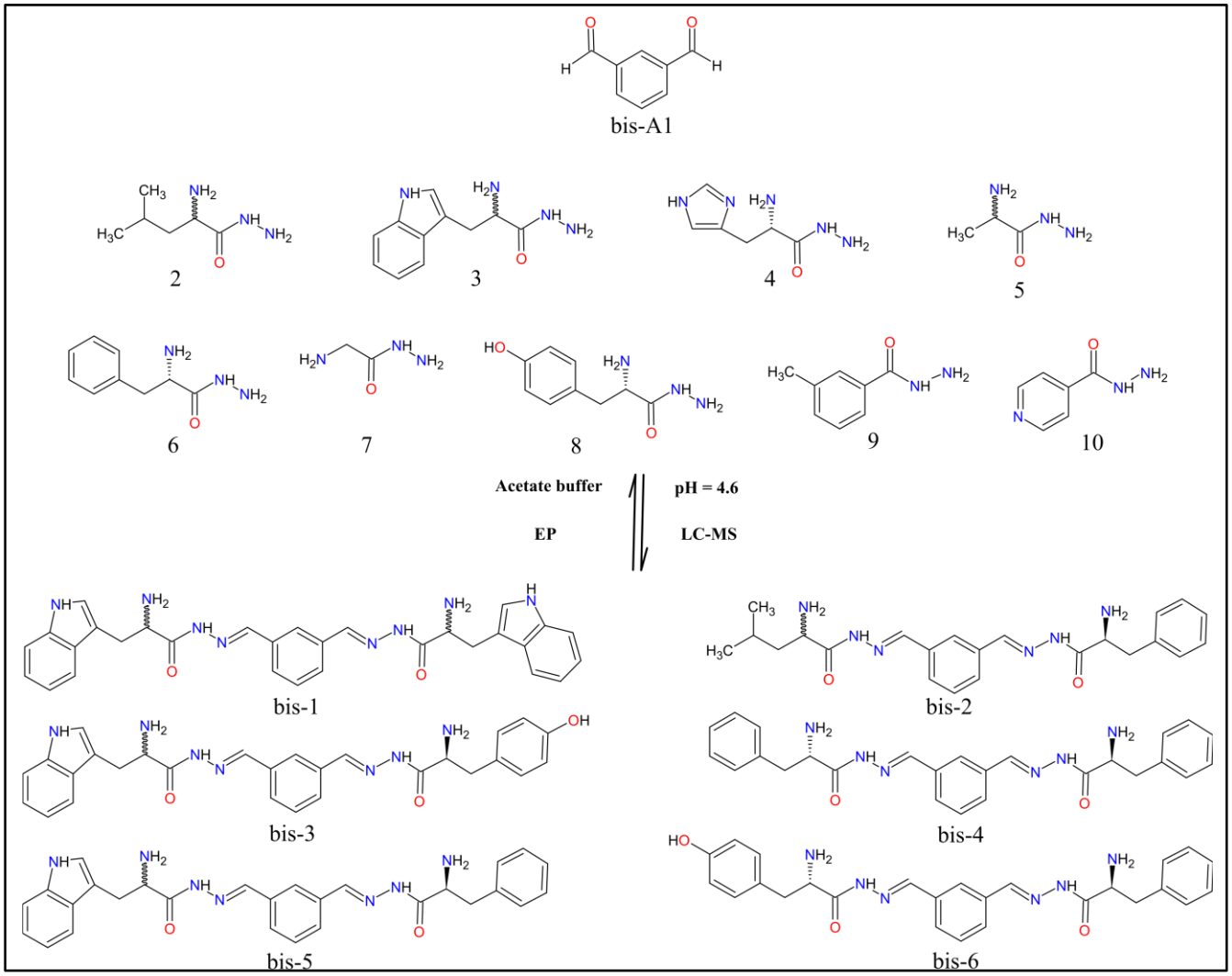


***Fig. 6.2**. Dynamic combinatorial library of bis-aldehyde and nine hydrazides linked together to bis-acylhydrazones.*

***Table 6.2**. Affinity data of the detected bis-acylhydrazones in contrast to the acylhydrazones from the first experiment.*

Inhibitors	IC_{50} (µM)	K_i (µM)	ΔG^0 (kJ·mol^{-1})[a]	LE (kJ·mol^{-1}·atom^{-1})
(*S*)-H4-A4	12.8 ± 0.4	6 ± 0.2	–30	0.27
(*R*)-H3-A5	14.5 ± 0.5	7 ± 0.2	–30	0.29
bis-1	0.054 ± 0.0005	0.0254 ± 0.0002	–49	0.29
bis-4	2.1 ± 0.1	0.98 ± 0.05	–34	0.25

[a] *The Gibbs free energy of binding was derived from the experimentally determined IC_{50} values.*

Remarkably, the bis-acylhydrazone bis-1 exhibited 240-fold improvement of potency compared to the previously identified acylhydrazones while keeping the ligand efficiency constant. We thus choose this ligand to validate the predicted binding mode and performed X-ray soaking analysis (e.g. 6.3 Results and Discussion).

6.3 Results and Discussion

The acylhydrazones (*S*)-H4-A4 and (*R*)-H3-A5 bind each with two molecules to the active site of EP (Fig. 6.3). As can be seen from Fig. 6.3, A, the acylhydrazone (*S*)-H4-A4 places its hydrophobic indolyl moiety into the S1 pocket, while establishing an H-bond interaction with Asp81 from the above located flap. Interestingly, it has already been reported that indol substituents prefer this binding region.[9] Furthermore, the mesityl moiety occupies the S2 binding pocket so that the ligand bridges between the S1 and S2 pockets with its acylhydrazone linker and the primary amino group is directly bound to the catalytic dyad. In contrast, the contact between the second promising (*R*)-H3-A5 and the catalytic aspartates is mediated by the catalytic water molecule (Fig. 6.3, B). This ligand also occupies the S1 pocket with its hydrophobic phenyl moiety, but the inverted stereochemistry at the introduced chiral center does not allow the same binding mode as discovered for (*S*)-H4-A4 to be adopted. Instead, the primary amino group is oriented away from the catalytic dyad as the acylhydrazone linker represents now the only possibility for an interaction with the dyad.

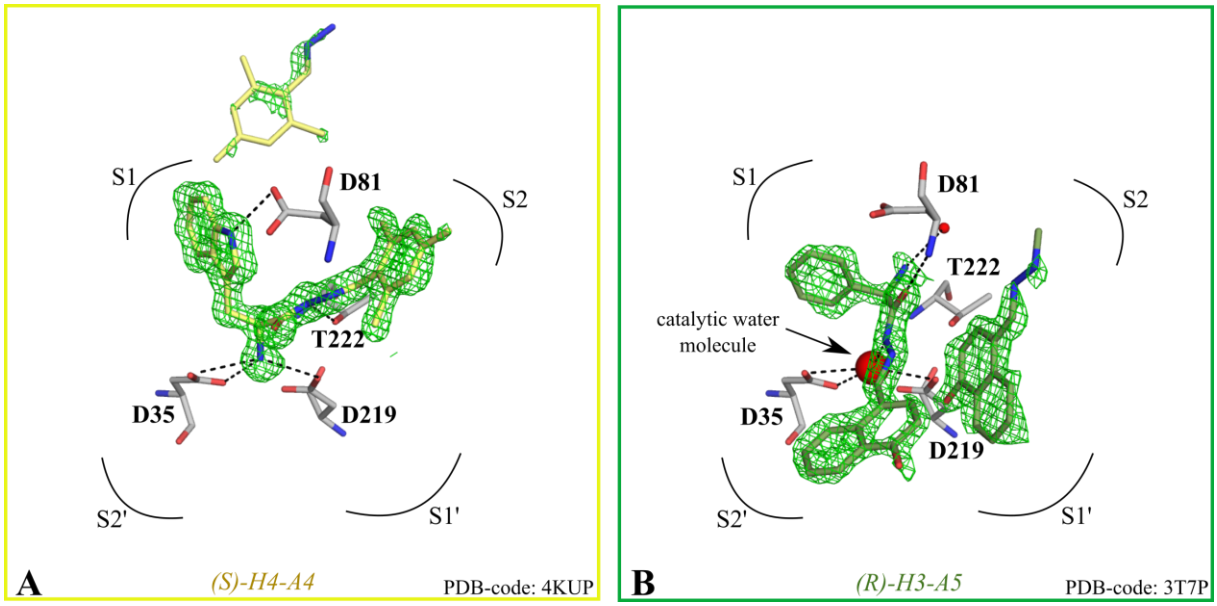


***Fig. 6.3**. Binding modes of the EP-acylhydrazone-complexes (S)-H4-A4 (A) and (R)-H3-A5 (B). The mF_o-DF_c density around the ligands is shown at 3σ.*

We also analyzed whether the acylhydrazone molecules, found in addition to the dyad binders in the crystal structures, were only observed due to the high ligand concentration applied during the soaking experiments (20 mM, e.g. Experimental Section). For this purpose, we performed additional soaking experiments with (*S*)-H4-A4 as we soaked the ligand once with five times

higher (100 mM, Fig. 6.4, A) and once with ten times lower concentrations (2 mM, Fig. 6.4, B). Remarkably, with the lower applied concentration only the molecule directly bound to the catalytic dyad could be assigned to the electron density, while the concentration increase to 100 mM led to the appearance of two additional molecules. However, due to the lack of additional interactions with EP we suppose that the appearance of those molecules is purely concentration induced.

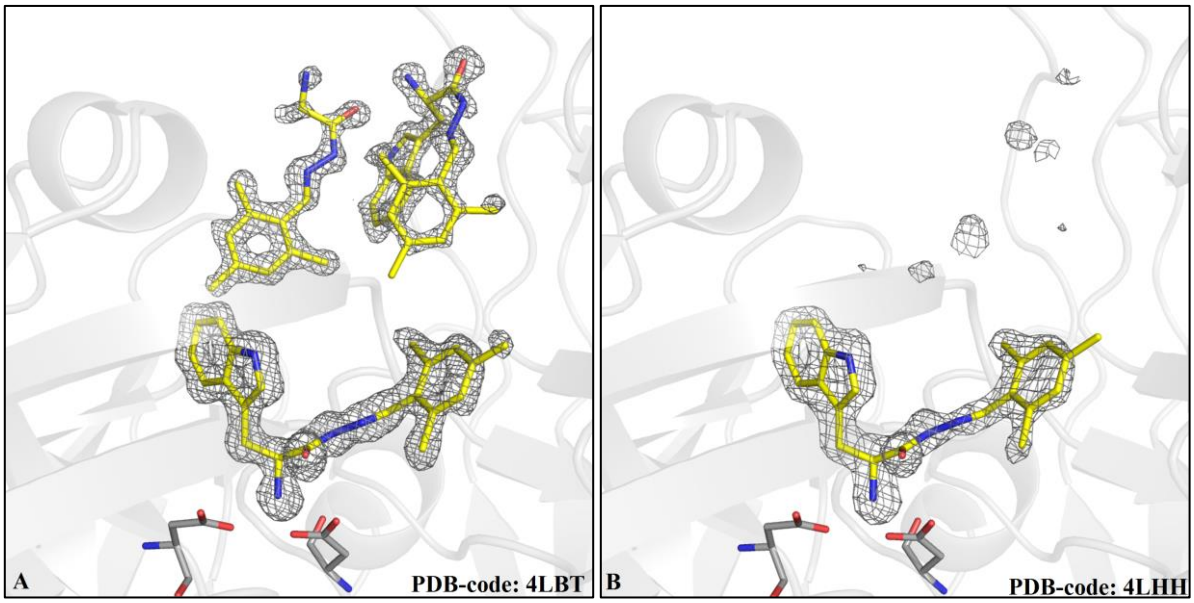


Fig.6.4. ***A****: (S)-H4-A4 at 100 mM concentration;* ***B****: (S)-H4-A4 at 2 mM concentration. The* mF_o-DF_c *density around the ligands is shown at 3σ.*

We also determined the crystal structure of the highly potent bis-acylhydrazone bis-1 bound to EP (Fig. 6.5, A). Interestingly enough, only part of the ligand could be assigned to the electron density. The visible part, which corresponds to (*S*)-H4-A4, used for the linking reaction, retained its original binding mode as observed before for the parent ligand. As a matter of fact, the second part of the bis-acylhydrazone could not be assigned to the electron density.

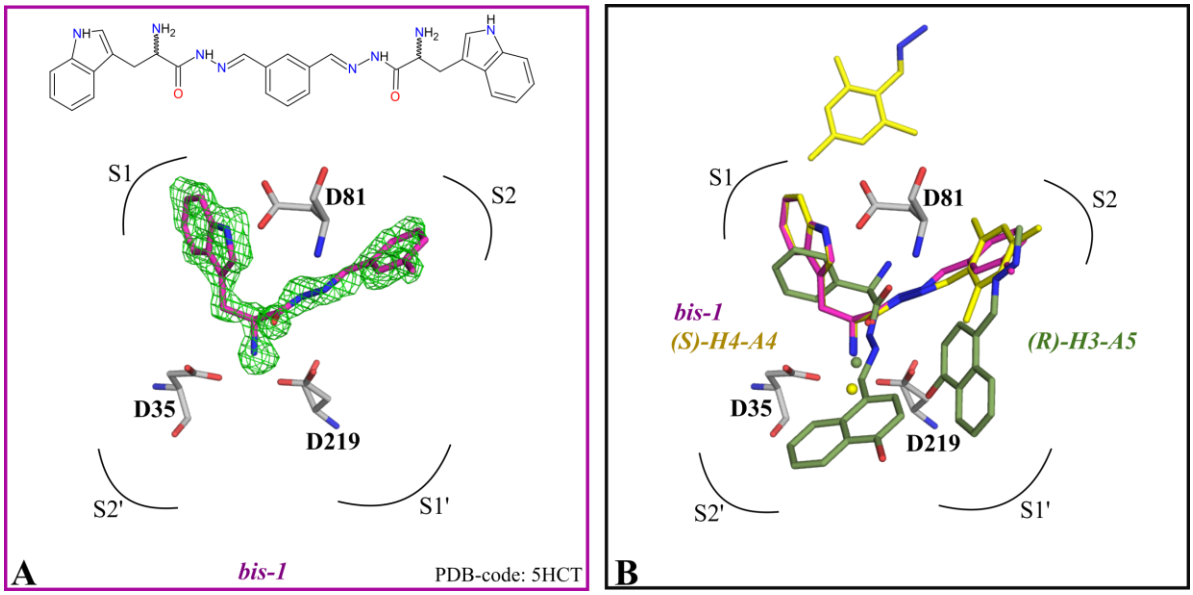


***Fig. 6.5**. A: Binding mode of the EP-bis-acylhydrazone-complex. **B**: Superimposition of all three analyzed ligands. The mF_o-DF_c density around the ligand in **A** is shown at 3σ.*

6.4 Conclusions

The application of dynamic combinatorial chemistry as method to generate compound libraries through self-assembly reactions has been successfully demonstrated in the above-described projects. The applicability of the method is highly interesting as by use of a few well-selected building blocks, hydrazides and aldehydes, a large number of potential high-affinity acyl- or bis-acylhydrazones are generated, covering a wide range of compounds as options to bind to the target protein requiring only a small amount of experimental effort. Furthermore, the method was successfully combined with de novo SBDD to identify the acylhydrazones with the highest affinity to EP. Moreover, the determined binding modes of the most promising acylhydrazones revealed alternative binding modes at the catalytic dyad in agreement with our fragment screening study (see chapters 3 and 4). The alternative binding modes were provoked by the inverted stereochemistry at the introduced chiral center, which led to the occupancy of different binding pockets. In a subsequent design step, the structural information has been further used to combine DCC with fragment linking, whereby the occupancy conservation of the already addressed pockets was intended. By this a bis-acylhydrazone could be obtained with a 240-fold improved potency compared to the parent single acylhydrazones.

Finally, the construction of specific receptors able to recognize particular molecules is a task that has been accomplished by biological systems over millions of years of evolution, and which is

hardly achievable via a straightforward synthetic concept. To reproduce evolution in a kind of test-tube approach, DCC appears ideally suitable as it can be applied to a wide range of target proteins and it has a great potential to rapidly identify potent hit molecules through natural selection by the target of interest.

6.5 Experimental Section

The synthesis of the hydrazides H1-H4 and hydrazides 2-7 along with the experimental procedures for the inhibitory assays, the modeling and docking studies, the STD-NMR spectra, and the UPLC chromatograms can be found in the associated papers (see 6.1 Introductory Remarks).

The data collection and refinement statistics of the acylhydrazone (*S*)-H4-A4 at the different concentrations along with the bis-acylhydrazone bis-1 are listed in Table 8.2 in the Appendix.

The EP crystallization procedure has already been described in chapter 3.7.1. For the soaking experiments, ligands (*S*)-H4-A4 at 2 mM (PDB-code: 4LHH), 20 mM (PDB-code: 4KUP), and 100 mM (PDB-code: 4LBT), (*R*)-H3-A5 at 20 mM (PDB-code: 3T7P), and bis-1 at 90 mM (PDB-code: 5HCT) were soaked for 48 h in a mixture of 75 mM NH_4Ac, 75 mM NaAc, 20% (w/v) PEG 4000, and 25% GOL before freezing in liquid nitrogen.

Datasets 4LHH, 4KUP, 4LBT, and 3T7P were collected at BESSY beamline 14.2 in Berlin and additionally processed with *HKL2000*.[10] Only dataset 5HCT was collected at BESSY beamline 14.1 and processed with *XDS*.[11] For all structures a simulated annealing was initially performed, followed by refinement of the *XYZ*-coordinates, individual B-factors, and occupancies performed with *Phenix*.[12] Structures with resolution better than 1.5 Å were refined anisotropically, with all other structures an isotropic refinement was performed. Model building was done with *Coot*.[13]

6.6 References

[1] Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. Dynamic combinatorial chemistry. *Drug Discovery Today* **2002**, *7*, 117–125.

[2] Lehn, J. M.; Eliseev, A. V. Dynamic combinatorial chemistry. *Science* **2001**, *291*, 2331–2332.

[3] Ramstrom, O.; Lehn, J.-M. Drug discovery by dynamic combinatorial libraries. *Nat. Rev. Drug Discovery* **2002**, *1*, 26–36.

[4] Mondal, M.; Radeva, N.; Koster, H.; Park, A.; Potamitis, C.; Zervou, M.; Klebe, G.; Hirsch, A. K. H. Structure-based design of inhibitors of the aspartic protease endothiapepsin by exploiting dynamic combinatorial chemistry. *Angew. Chem., Int. Ed. Engl.* **2014**, *53*, 3259–3263.

[5] Gerber, P. R.; Muller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 251–268.

[6] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

[7] Wang, Y.-S.; Liu, D.; Wyss, D. F. Competition STD NMR for the detection of high-affinity ligands and NMR-based screening. *Magn. Reson. Chem.* **2004**, *42*, 485–489.

[8] Mondal, M.; Radeva, N.; Fanlo-Virgos, H.; Otto, S.; Klebe, G.; Hirsch, A. K. H. Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin: Fragment-Based Drug Design Facilitated by Dynamic Combinatorial Chemistry. *Angew. Chem., Int. Ed. Engl.* **2016**, *55*, 9422–9426.

[9] Kuhnert, M.; Koster, H.; Bartholomaus, R.; Park, A. Y.; Shahim, A.; Heine, A.; Steuber, H.; Klebe, G.; Diederich, W. E. Tracing binding modes in hit-to-lead optimization: chameleon-like poses of aspartic protease inhibitors. *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 2849–2853.

[10] Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **1997**, 307–326.

[11] Krug, M.; Weiss, M. S.; Heinemann, U.; Mueller, U. XDSAPP: A graphical user interface for the convenient processing of diffraction data using XDS. *J. Appl. Crystallogr.* **2012**, *45*, 568–572.

[12] Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213–221.

[13] Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.

7. Summary

The present work focuses on the evaluation and comparison of different fragment-based approaches, for which purpose the model system Endothiapepsin (EP) has been used. The enzyme belongs to the large family of aspartic proteases. Many representatives are involved in the pathogenesis of serious diseases such as malaria, Alzheimer's disease, AIDS, hypertension, and fungal infections. Due to the rapid resistance development of those targets against commercially available drugs, the compliance issues due to side effects, and the high therapy costs, the development of novel drugs is highly important.

We dealt with fragments as these are optimally suitable for optimization strategies due to their low-molecular weight and high ligand efficiency. For this purpose, we used a 361-entry in-house fragment library compiled based on physico-chemical properties similar to the rule-of-three. In cooperation with partners from industry we applied six different techniques to screen the fragment library: saturation-transfer difference NMR (STD-NMR), thermal shift assay (TSA), reporter-displacement assay (RDA), microscale thermophoresis (MST), fluorescence-based biochemical assay (HCS), and native mass spectrometry (MS). However, due to method's diversity and their individual sensitivities for detecting weak binders, varying fragment concentrations at applied assay conditions, as similar as possible, were used (chapter 2, Table 2.1). For each assay a threshold was set based on the assay systems used and on the practical experience given in each of the individual contributing groups. This might have influenced the screening results. Moreover, as not all fragments could be measured, mainly due to solubility issues, the relative hit rates normalized by the total number of fragments used in the individual assays were also considered. Altogether, the RDA and biochemical assay showed comparable number of hits (50 versus 56). However, the number of hits is drastically lower for the TSA and MS assays (25 and 8, respectively), although the total number of fragments used for each of the four assays was very similar (325, 358, 330, and 342, respectively). Alarmingly, the overlap between the hits produced by all six methods was very low. While only 41 out of all 361 fragment library entries could be identified by at least two methods, only 3 were identified by five methods, and not a single fragment was identified as hit by all six methods taken together (chapter 2, Fig. 2.1).

We thus decided to perform X-ray crystallography with individual fragments separately soaked into beforehand prepared *apo* EP crystals. For this purpose very high fragment concentrations of 90 mM, hardly applicable in any of the other screening techniques, were used. Intriguingly, we were able to identify 71 fragment hits as bound to EP, corresponding to a 20% hit rate. Perturbingly, only 30% of the 71 hits were predicted by only one of the beforehand applied screening techniques, clearly emphasizing that any screening strategy comprising at least two prescreening methods would have been able to identify only 19 (27%) of the 71 crystallographic hits. This implicates that we would have failed to identify 73% of the crystallographically detected fragment hits. As the fragment hits were spread throughout the entire EP binding cleft with few examples found on the protein surface, we divided the hits into two groups: catalytic dyad binders and remote binders.

Chapter 3 deals with the catalytic dyad binders. Therein we describe a variety of warheads, which address the catalytic aspartates either directly or via the catalytic water molecule W501, while occupying the S1 and S1' binding pockets. Amazingly, we were able to discover novel warheads, which have not been reported as potential dyad binders to any other relevant aspartic protease yet. One of these is the group of hydrazides, four of which occupy the S1' pocket, while one (**F109**) is oriented oppositely, accommodated in the S1 pocket. After additionally performed docking analysis, soaking experiments of a similar fragment, and ITC measurements to investigate this behavior, we found that the displacement of a single conserved water molecule was responsible for the swapped orientation of **F109** and the related affinity enhancement. Furthermore, we found fragments, which occupied both pockets simultaneously (**F075**, **F114**, **F162**, **F255**, and **F260**), offering an optimal platform for further optimization strategies into adjacent pockets on each side of the catalytic aspartates.

The very useful information regarding preferred spots of binding remote from the catalytic dyad is described in detail in chapter 4. The so called hot spots represent key pocket residues, which have to be addressed in an optimization procedure in order to achieve optimal ligand binding. For example, many large and hydrophobic fragments addressed the hot spot residue Phe291 in the S6 binding pocket. Moreover, the moieties performing the contact with this hot spot exhibited diverse chemistry, allowing a conclusive mapping of a broader part of the very large chemical space. Nevertheless, the position of Phe291 can vary within the different members of the pepsin-like aspartic protease family. For instance, in renin an aspartate is found at the same place, which

would require a Lipinski donor instead of a hydrophobic substituent. Such differences, however, can be very well used with regard to selectivity issues. Furthermore, we discovered fragments, which occupy simultaneously the adjacent S1 and S3 and the S4 and S6 binding pockets, suggesting valuable ideas on how to merge these pockets while reducing the size of peptidomimetic inhibitors.

The high-throughput potential of X-ray crystallography (chapters 3 and 4) compared to the other six prescreening techniques presented in chapter 2 is still very low. Because of this, a lot of effort has been invested in academia and industry to extend the method's applicability as a primary screening technique. Whenever used as such, practitioners exposed protein crystals to a mixture of compounds to accelerate the hit identification step. However, we cannot recommend the use of cocktail experiments with clear conscience as we believe that parameters such as compound solubility, chemical reactivity, competitive binding poses, and crystal damage, often resulting in reduced ligand occupancy, diffuse electron density, or deviating fragment binding poses are provoked by the presence of multiple compounds in a mixture. These issues are exemplified in chapter 5 where we directly compared the difference electron density in the binding pocket of EP between singly soaked fragments and cocktails prepared from those. For example, many of the fragments used in a mixture of two failed to be identified as hits, whereas clear density indicating these fragments could be observed in single soaking experiments. Moreover, the clear assignment of a fragment to diffuse electron density usually requires additional experiments, which confirm the binding pose. When considering the elaborate steps, even more time has to be spent, which extremely slows down the hit identification procedure.

The only plausible reason to use several compounds in a mixture is when a reaction between those is aimed. This is an approach named dynamic combinatorial chemistry (DCC) and described in chapter 6. The compounds, hydrazides and aldehydes, chosen for the mixtures react with one another involving a dehydration step to merge to larger acylhydrazones. The formation of the strongest binders among others, acylhydrazones (*S*)-H4-A4 and (*R*)-H3-A5, was induced by natural selection as the target protein EP extracts by potent binding such compounds from the equilibrium mixture. Moreover, in a follow-up optimization project, which started based on the binding modes of the two acylhydrazones, the combination of a bis-aldehyde with hydrazides resulted in the natural selection of a bis-acylhydrazone by EP, with an 240-fold improved potency compared to the initial acylhydrazones, resulting in a two-digit nanomolar lead.

7. Zusammenfassung

Die vorgelegte Arbeit beruht auf der Auswertung und dem Vergleich verschiedener, fragmentbasierter Verfahren, für welche Zwecke das Modellprotein Endothiapepsin (EP) verwendet wurde. Das Enzym gehört zu der großen Familie der Aspartylproteasen. Viele Vertreter dieser Familie sind in die Pathogenese schwerer Krankheiten wie Malaria, Alzheimer, AIDS, Bluthochdruck, und Pilzinfektionen involviert. Aufgrund der raschen Resistenzentwicklung dieser Enzyme gegen kommerziell verfügbare Medikamente, der mangelnden Patientencompliance aufgrund von unerwünschten Nebenwirkungen und hoher Therapiekosten, ist die Entwicklung neuartiger Medikamente besonders wichtig.

Fragmente wurden als Ausgangsbasis gewählt, da sich diese aufgrund ihres niedrigen Molekulargewichtes und ihrer hohen Ligandeffizienz hervorragend als Startpunkt für Optimierungsstrategien eignen. Für diese Zwecke wurde eine interne Fragmentbibliothek verwendet, die 361 Fragmente enthält und die auf der Grundlage von physikalisch-chemischen Eigenschaften ähnlich der Astex rule-of-three basiert. Um die interne Fragmentbibliothek zu durchmustern, wurden zusammen mit Kooperationspartnern aus der Industrie sechs verschiedene Screeningmethoden angewendet: Sättigungstransferdifferenz mit Kernspinresonanz (STD-NMR), Thermal Shift Assay (TSA), Reporterverdrängungsassay (RDA), Thermophorese (MST), ein fluoreszenzbasierter biochemischer Assay (HCS), und Massenspektrometrie (MS). Aufgrund der Methodenvielfalt und der individuell unterschiedlichen Sensitivität schwache Binder zu detektieren, wurden unterschiedliche Fragmentkonzentrationen unter möglichst ähnlich gehaltenen Assaybedingungen verwendet (Kapitel 2, Tabelle 2.1). Für jeden Assay wurde ein Schwellenwert festgesetzt, abhängig von den angewendeten Assaysystemen und der praktischen Erfahrung jeder der involvierten Arbeitsgruppen. Darüber hinaus wurden die relativen Hitraten, normiert auf die Gesamtzahl aller in dem jeweiligen Assay verwendeten Fragmente, berücksichtigt, da aufgrund von Löslichkeitsproblemen nicht immer die gesamte Fragmentbibliothek vermessen werden konnte. Im Allgemeinen haben die RDA und HCS Assays eine vergleichbare Anzahl an Hits gezeigt (50 (RDA) gegenüber 56 (HCS)). Jedoch ist die Anzahl der Hits für die TSA und MS Assays drastisch geringer (25 (TSA) gegenüber 8 (MS)), obwohl die Gesamtzahl der verwendeten Fragmente in allen vier Assays sehr ähnlich war (325 (RDA), 358 (HCS), 330 (TSA), und 342 (MS)). Auffällig ist allerdings die geringe Überlappung

der Hits, die von jeder der sechs Methoden detektiert wurden. Während nur 41 von 361 Fragmenten in der Bibliothekdatenbank von mindestens zwei Methoden detektiert werden konnten, waren es nur drei, die von fünf Methoden und kein einziges Fragment, welches von allen Methoden gleichzeitig detektiert wurde (Kapitel 2, Abb. 2.1).

Aus diesem Grund wurde jedes der 361 Fragmente einzeln in *apo* EP-Kristalle diffundiert um mit den hieraus erhaltenen Daten eine Röntgenstrukturanalyse durchzuführen. Hierfür wurde eine sehr hohe Fragmentkonzentration von 90 mM verwendet, welche für andere Screeningmethoden kaum anwendbar ist. Interessanterweise konnten 71 gebundene Fragmente detektiert werden, was einer Hitrate von 20% entspricht. Auffallend war allerdings die Tatsache, dass nur 30% der 71 Hits von maximal einer der sechs zuvor durchgeführten Screeningmethoden detektiert werden konnten. Dies stellt deutlich heraus, dass jede Screeningkaskade, bestehend aus mindestens zwei Methoden, nur 19 (27%) der 71 kristallographischen Hits detektiert hätte. Dies hätte zur Folge gehabt, dass 73% der 71 kristallographischen Hits nicht als solche erkannt worden wären. Da die bindenden Fragmente, abgesehen von wenigen Beispielen an der Proteinoberfläche, über die gesamte Bindetasche von EP verteilt waren, wurden die Hits in zwei Kategorien aufgeteilt: direkt- und fernbindende Fragmente.

In Kapitel 3 wurden die Direktbinder behandelt. Dort wurde eine Vielzahl von Kopfgruppen beschrieben, die die katalytische Dyade entweder direkt oder über das katalytische Wasser W501 adressieren, während sie die S1 und S1‘ Bindetaschen besetzen. Überraschenderweise konnten neuartige, bis jetzt noch nie zuvor als potentielle Binder für andere relevante Aspartylproteasen beschriebene Kopfgruppen identifiziert werden. Eine davon ist die Gruppe der Hydrazide. Vier davon besetzen die S1‘ Tasche, während ein Fragment (**F109**) sich in die entgegengesetzte Richtung orientiert und stattdessen die S1 Tasche besetzt. Nach anschließend durchgeführten Dockinganalysen, Soakingexperimenten mit einem Fragmentanalogon, und isothermaler Titrationskalorimetriemessungen wurde ersichtlich, dass die Verdrängung eines einzelnen konservierten Wassermoleküls für die getauschte Orientierung und die damit verbundene Affinitätserhöhung von **F109** verantwortlich war. Darüber hinaus wurden Fragmente gefunden, die beide Bindetaschen gleichzeitig besetzen (**F075**, **F114**, **F162**, **F255**, und **F260**), was eine optimale Grundlage für Optimierungsstrategien in die benachbarten Taschen auf jeder Seite der katalytischen Aspartate bietet.

Bindestellen abseits der katalytischen Dyade wurden in Kapitel 4 beschrieben. Die so genannten „*hot-spots*“ stellen wichtige Aminosäuren dar, die in einem Optimierungsprozess adressiert werden sollen, um optimale Ligandbindung zu erzielen. Zum Beispiel haben viele große und hydrophobe Fragmente den *hot-spot* Rest Phe291 in der S6 Tasche adressiert. Darüber hinaus haben die Substituenten der Fragmente, die diesen Rest adressieren, unterschiedlichste chemische Beschaffenheit, was eine schlüssige Zuordnung eines größeren Teils des riesigen chemischen Raumes erlaubt. Dennoch kann die Position von Phe291 in anderen Proteinen aus der Familie der pepsinähnlichen Aspartylproteasen variieren. Zum Beispiel findet sich in Renin an dieser Stelle ein Aspartatrest, welches einen Lipinski Donor anstatt von einem hydrophoben Substituenten verlangt. Jedoch können solche Unterschiede in Bezug auf Selektivität wunderbar genutzt werden. Zusätzlich wurden Fragmente identifiziert, die gleichzeitig die S1 und S3 sowie die S4 und S6 Bindetaschen besetzten. Diese deuten auf die hervorragende Optimierungsmöglichkeit hin, bei großen peptidomimetischen Inhibitoren eine Reduzierung der Größe zu erreichen, indem die jeweiligen Bindetaschen zu einer größeren verschmolzen werden könnten.

Das Durchsatzpotenzial der Röntgenstrukturkristallographie (Kapitel 3 und 4) verglichen mit den anderen sechs in Kapitel 2 beschriebenen Screeningverfahren, ist noch sehr gering. Aus diesem Grund wurde enormer Aufwand in akademischen Kreisen und der Industrie betrieben um die Anwendbarkeit der Röntgenstrukturanalyse als primäre Screeningmethode zu erweitern. Wann immer diese als solche genutzt wurde, wurde eine Mischung aus Verbindungen in Proteinkristalle diffundiert um die Hitidentifizierung zu beschleunigen. Dieses Verfahren können wir jedoch nicht empfehlen, da wir glauben, dass Parameter wie Löslichkeit, Reaktivität zwischen den Verbindungen, kompetitive Bindungsposen, und Kristallschäden, die oft zu reduzierter Ligandbesetzung, diffuser Elektronendichte oder unterschiedlichen Fragmentbindeposen führen, durch die Mehrzahl an Verbindungen in einer Mischung stark beeinflusst werden. Probleme solcher Art sind in Kapitel 5 dokumentiert, in dem die Differenzelektronendichte zwischen einzeln gesoakten Fragmenten und Cocktails aus diesen direkt miteinander verglichen werden. Zum Beispiel konnten viele der Fragmente, die in zweier Cocktails verwendet wurden, nicht als Hits identifiziert werden, wohingegen in den Einzelsoaks deutliche Dichte, die das gebundene Fragment indiziert, beobachtet werden konnte. Darüber hinaus erfordert die klare Zuordnung eines Fragmentes zu diffuser Elektronendichte zusätzliche Experimente, die den Bindungsmodus

bestätigen. Durch die Berücksichtigung solcher aufwendigen Schritte wird die Verwendung dieser Methode noch zeitintensiver und verlangsamt die Hitidentifizierungsprozedur enorm.

Der einzig plausible Grund für die Verwendung mehrerer Verbindungen in einem Cocktail ist eine beabsichtigte Reaktion zwischen einzelnen Komponenten. Dieses Verfahren wird dynamische kombinatorische Chemie (DCC) genannt und in Kapitel 6 beschrieben. Als Verbindungen für die Mischung wurden Hydrazide und Aldehyde gewählt, die in einer Kondensationsreaktion zu Acylhydrazonen reagieren. Die Bildung der stärksten Liganden, die Acylhydrazone (*S*)-H4-A4 und (*R*)-H3-A5, wurde durch natürliche Selektion veranlasst, da das Zielprotein EP solche potenten Binder aus dem Mischungsgleichgewicht zieht. In einer Folgestudie erfolgte basierend auf den Bindungsmodi der zwei Acylhydrazone die Kombination von einem bis-Aldehyd mit Hydraziden und hieraus die natürliche Selektion eines bis-Acylhydrazones durch EP. Das auf diese Weise identifizierte bis-Acylhydrazon wies eine 240-fach höhere Affinität im Vergleich zu den ursprünglichen Acylhydrazonen auf, resultierend in einem zweistellig nanomolaren Binder.

8. Appendix

8.1 Data collection and refinement statistics of fragments

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 1***

Fragment number	**004**	**005**	**014**	**017**
PDB code	**4Y36**	**4Y3E**	**4Y4U**	**4Y3P**
Data collection and processing				
Wavelength (Å)	0.91842	0.91841	0.91841	0.91841
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
Detector	MX-225 CCD	MX-225 CCD	MX-225 CCD	MX-225 CCD
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.5, 73.0, 52.6	45.3, 73.0, 52.8	45.6, 72.8, 52.7	45.4, 72.7, 52.5
α, β, γ (°)	90.0, 108.9, 90.0	90.0, 109.5, 90.0	90.0, 108.9, 90.0	90.0, 109.0, 90.0
Resolution range (Å)	43.0-1.59 (1.69-1.59)	42.7-1.25 (1.33-1.25)	50.0-1.75 (1.78-1.75)	42.9-1.55 (1.64-1.55)
Wilson B factor ($Å^2$)	12.5	10.2	17.1	12.5
No. of unique reflections	43597 (7012)	85993 (13556)	31788 (1542)	46729 (7195)
Multiplicity	4.1 (4.2)	4.3 (4.2)	4.0 (4.0)	4.2 (4.1)
R_{sym} (%)	7.3 (49.1)	4.4 (48.1)	6.8 (46.7)	6.3 (50.8)
Completeness (%)	99.4 (98.9)	96.2 (94.0)	97.8 (95.9)	98.9 (94.5)
$\langle I/\sigma(I) \rangle$	14.6 (2.9)	20.3 (3.2)	17.3 (3.2)	17.1 (2.8)
Refinement				
Resolution range	41.1 - 1.59	41.1 - 1.25	43.2 - 1.75	42.9 - 1.55
No. of reflections used for refinement (work (free))	41412 (2185)	81693 (4300)	30158 (1607)	44392 (2337)
R_{cryst} (%)	17.2	11.6	15.5	11.8
R_{free} (%)	19.9	13.5	18.9	16.1
No. of refined residues	330	330	330	330
No. of fragment atoms	11	10	13	24
No. of other ligand atoms	28	30	23	12
No. of water molecules	310	342	321	317
RMSD, bond length (Å)	0.009	0.006	0.006	0.008
RMSD, bond angles (°)	1.2	1.2	1.1	1.1
Ramachandran plot (%) [b]				
Most favoured	93.5	93.9	93.1	93.5
Additionally allowed	6.5	6.1	6.9	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	13.5	12.1	18.6	13.9
Main chain	13.3	11.2	18.2	13.0
Side chain	13.8	12.9	19.2	14.8
Fragment atoms	27.8	17.6	27.5	36.8
Other ligand atoms [d]	33.5	36.6	39.6	27.8
Waters	27.1	29.3	29.9	30.8

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 2***

Fragment number	**031**	**034**	**035**	**041**
PDB code	**4Y5C**	**5DQ1**	**4Y3W**	**4Y3Z**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.89500
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1	BESSY BL14.3
Detector	MX-225 CCD	MX-225 CCD	MX-225 CCD	Mar CCD 165
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.1, 52.7	45.4, 73.3, 53.0	45.1, 73.6, 52.4	45.2, 73.1, 52.7
α, β, γ (°)	90.0, 109.4, 90.0	90.0, 109.8, 90.0	90.0, 109.2, 90.0	90.0, 109.4, 90.0
Resolution range (Å)	42.8-1.33 (1.41-1.33)	42.8-1.49 (1.58-1.49)	42.6-1.58 (1.67-1.58)	42.6-1.12 (1.19-1.12)
Wilson B factor ($Å^2$)	10.0	10.8	13.9	9.5
No. of unique reflections	68918 (7630)	53305 (8493)	44277 (7076)	122811 (18556)
Multiplicity	3.8 (2.5)	4.1 (3.9)	4.2 (4.1)	4.0 (3.4)
R_{sym} (%)	4.5 (37.6)	7.4 (46.2)	6.2 (50.2)	5.5 (42.7)
Completeness (%)	92.3 (63.3)	99.5 (98.5)	99.6 (98.6)	98.5 (92.3)
$<I/\sigma(I)>$	18.9 (2.4)	14.1 (2.9)	16.2 (2.8)	14.3 (2.8)
Refinement				
Resolution range	42.8 - 1.33	41.2 - 1.49	36.8 - 1.58	42.7 - 1.12
No. of reflections used for refinement (work (free))	65492 (3426)	50639 (2666)	42063 (2214)	116670 (6141)
R_{cryst} (%)	11.9	13.4	12.3	10.9
R_{free} (%)	14.6	17.5	16.3	12.5
No. of refined residues	330	330	330	330
No. of fragment atoms	8	18	24	19
No. of other ligand atoms	39	10	10	34
No. of water molecules	301	336	302	328
RMSD, bond length (Å)	0.007	0.008	0.005	0.007
RMSD, bond angles (°)	1.2	1.2	1.0	1.3
Ramachandran plot (%) [b]				
Most favoured	93.5	93.5	93.5	93.9
Additionally allowed	6.5	6.5	6.5	6.1
Generously allowed	0.0	0.0	0.0	0
Disallowed	0.0	0.0	0.0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	11.6	12.2	16.8	11.7
Main chain	10.7	11.3	15.7	10.8
Side chain	12.5	13.1	17.7	12.7
Fragment atoms	41.3	27.9	34.8	21.8
Other ligand atoms [d]	36.4	24.7	25.3	29.9
Waters	**031**	26.4	31.2	26.9

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 3***

Fragment number	**042**	**048**	**051**	**052**
PDB code	**4Y43**	**5DQ2**	**4Y4X**	**4Y51**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.89120	0.91841
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.3	BESSY BL14.1
Detector	MX-225 CCD	MX-225 CCD	Mar CCD 165	MX-225 CCD
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.4, 72.4, 52.5	45.6, 72.4, 52.7	45.1, 72.8, 52.5	45.1, 72.9, 52.5
α, β, γ (°)	90.0, 108.7, 90.0	90, 108.8, 90	90.0, 109.5, 90.0	90.0, 109.5, 90.0
Resolution range (Å)	43.0-1.48 (1.57-1.48)	43.1-1.51 (1.61-1.51)	42.5-1.67 (1.77-1.67)	42.5-1.60 (1.70-1.60)
Wilson B factor ($Å^2$)	10.9	12.9	10.4	10.6
No. of unique reflections	52124 (7097)	48966 (6641)	37024 (5927)	40865 (5688)
Multiplicity	4.2 (4.0)	4.2 (4.1)	4.2 (4.1)	4.1 (3.4)
R_{sym} (%)	4.8 (40.3)	5.6 (50.3)	9.5 (47.8)	6.7 (63.2)
Completeness (%)	96.3 (81.2)	96.9 (82.1)	99.3 (99.0)	97.1 (84.0)
$\langle I/\sigma(I) \rangle$	21.6 (3.3)	19.0 (3.0)	15.3 (3.4)	17.3 (3.1)
Refinement				
Resolution range	41.0 - 1.48	43.1 - 1.51	42.5-1.67	36.5 - 1.60
No. of reflections used for refinement (work (free))	49517 (2607)	46517 (2449)	35170 (1854)	38819 (2046)
R_{cryst} (%)	12.6	12.1	14.9	17.0
R_{free} (%)	16.7	16.4	17.8	18.9
No. of refined residues	330	330	330	330
No. of fragment atoms	15	17	18	11
No. of other ligand atoms	32	16	30	10
No. of water molecules	339	309	292	268
RMSD, bond length (Å)	0.005	0.008	0.006	0.006
RMSD, bond angles (°)	1.0	1.2	1.0	1.0
Ramachandran plot (%) [b]				
Most favoured	93.5	93.5	93.5	93.1
Additionally allowed	6.5	6.5	6.5	6.9
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	12.2	14.5	11.8	12.6
Main chain	11.8	13.6	11.2	12.1
Side chain	12.7	15.4	12.6	13.0
Fragment atoms	22.2	34.5	18.1	21.3
Other ligand atoms [d]	28.3	32.7	28.1	24.6
Waters	29.1	29.1	23.7	24.7

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 4***

Fragment number	**F analogue of 052**	**054**	**056**	**058**
PDB code	**5ISK**	**4Y53**	**5DQ2**	**4Y56**
Data collection and processing				
Wavelength (Å)	1.00000	0.91841	0.91841	0.91841
Beamline	ELETTRA BL 5.2R	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
Detector	PILATUS 2M	MX-225 CCD	MX-225 CCD	MX-225 CCD
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.1, 72.9, 52.7	45.2, 72.6, 52.7	45.6, 72.4, 52.7	45.2, 72.8, 52.5
α, β, γ (°)	90.0, 109.3, 90.0	90.0, 109.8, 90.0	90, 108.8, 90	90.0, 109.7, 90.0
Resolution range (Å)	42.6-1.14 (1.21-1.14)	42.5-1.62 (1.72-1.62)	43.1-1.51 (1.61-1.51)	42.5-1.63 (1.73-1.63)
Wilson B factor ($Å^2$)	12.0	10.9	12.9	10.9
No. of unique reflections	116418 (18671)	40417 (6365)	40942 (6168)	39838 (6417)
Multiplicity	3.3 (3.1)	4.1 (3.7)	4.2 (4.1)	4.2 (4.0)
R_{sym} (%)	3.3 (48.8)	7.6 (42.8)	5.6 (50.3)	5.7 (31.7)
Completeness (%)	99.5 (99.1)	99.0 (97.3)	96.9 (82.1)	99.4 (99.2)
$\langle I/\sigma(I) \rangle$	16.8 (2.5)	14.5 (3.2)	19.0 (3.0)	20.1 (4.4)
Refinement				
Resolution range	42.6 - 1.14	41.0 - 1.62	43.1 - 1.51	42.5 - 1.63
No. of reflections used for refinement (work (free))	110572 (5814)	38390 (2027)	38897 (2045)	37846 (1992)
R_{cryst} (%)	11.9	18.5	12.1	14.3
R_{free} (%)	14.6	20.6	16.4	17.1
No. of refined residues	330	330	330	330
No. of fragment atoms	11	34	17	9
No. of other ligand atoms	55	10	16	22
No. of water molecules	344	295	309	359
RMSD, bond length (Å)	0.007	0.007	0.008	0.01
RMSD, bond angles (°)	1.2	1.0	1.2	1.3
Ramachandran plot (%) [b]				
Most favoured	93.5	93.1	93.5	93.5
Additionally allowed	6.5	6.9	6.5	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	14.5	12.0	14.5	11.4
Main chain	13.6	11.7	13.6	11.1
Side chain	15.3	12.3	15.4	11.9
Fragment atoms	22.7	22.8	34.5	17.8
Other ligand atoms [d]	33.5	22.0	32.7	25.9
Waters	32.9	23.0	29.1	25.5

[a] Statistic for the highest-resolution shell are shown in parenthesis; [b] Calculated by PROCHECK; [c] Calculated by MOLEMAN; [d] Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 5***

Fragment number	**063**	**066**	**073**	**075**
PDB code	**4Y57**	**5DQ4**	**4Y4Z**	**4Y39**
Data collection and processing				
Wavelength (Å)	0.91841	0.8912	0.91841	0.8944
Beamline	BESSY BL14.1	BESSY BL14.3	BESSY BL14.1	BESSY BL14.3
Detector	MX-225 CCD	Mar CCD 165	MX-225 CCD	Mar CCD 165
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.6, 73.3, 53.1	45.1, 72.8, 52.6	45.3, 73.4, 52.9	45.2, 73.0, 52.6
α, β, γ (°)	90.0, 109.9, 90.0	90, 109.7, 90	90.0, 109.5, 90.0	90.0, 109.5, 90.0
Resolution range (Å)	42.8-1.49 (1.58-1.49)	42.5-1.15 (1.22-1.15)	42.7-1.48 (1.57-1.48)	42.7-1.20 (1.27-1.20)
Wilson B factor ($Å^2$)	10.9	9.0	10.7	9.4
No. of unique reflections	53579 (8621)	111456 (16822)	54295 (8737)	99993 (15991)
Multiplicity	4.2 (4.2)	4.1 (3.7)	4.2 (4.1)	4.2 (4.1)
R_{sym} (%)	7.5 (49.4)	3.9 (26.1)	6.6 (49.2)	4.9 (47.5)
Completeness (%)	99.8 (99.6)	98.2 (91.8)	99.8 (99.5)	99.4 (98.7)
$\langle I/\sigma(I) \rangle$	13.5 (2.9)	21.4 (5.1)	16.6 (3.0)	17.0 (3.0)
Refinement				
Resolution range	41.2 - 1.49	16.5 - 1.15	34.9 - 1.48	28.0 - 1.20
No. of reflections used for refinement (work (free))	50900 (2679)	105898 (5558)	51580 (2713)	94993 (5000)
R_{cryst} (%)	11.7	11.1	12.3	11.7
R_{free} (%)	15.3	13.0	15.7	13.9
No. of refined residues	330	330	330	330
No. of fragment atoms	15	13	8	16
No. of other ligand atoms	20	6	30	33
No. of water molecules	326	339	261	272
RMSD, bond length (Å)	0.008	0.008	0.005	0.008
RMSD, bond angles (°)	1.2	1.3	1.1	1.3
Ramachandran plot (%) [b]				
Most favoured	93.5	93.5	93.1	94.2
Additionally allowed	6.5	6.5	6.9	5.8
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	11.8	10.9	11.6	11.8
Main chain	10.9	10.1	10.6	10.7
Side chain	12.6	11.7	12.4	12.7
Fragment atoms	13.2	22.4	27.9	12.6
Other ligand atoms [d]	24.6	12.9	29.6	27,1
Waters	27.2	26.5	24.7	25.5

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 6***

Fragment number	**078**	**081**	**103**	**109**
PDB code	**4Y3F**	**4Y50**	**4Y3M**	**4Y3Q**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.2	BESSY BL14.2
Detector	MX-225 CCD	MX-225 CCD	MX-225 CCD	MX-225 CCD
Space group	$P2_1$	$P2_1$	$P2_12_12_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.4, 73.0, 52.8	45.3, 73.2, 52.7	45.2, 72.5, 104.1	45.3, 72.9, 52.6
α, β, γ (°)	90.0, 109.8, 90.0	90.0, 109.4, 90.0	90.0, 90.0, 90.0	90.0, 109.3, 90.0
Resolution range (Å)	42.7-1.40 (1.48-1.40)	42.7-1.32 (1.40-1.32)	45.2-1.55 (1.64-1.55)	42.7-1.17 (1.24-1.17)
Wilson B factor ($Å^2$)	8.2	9.7	11.1	9.1
No. of unique reflections	63130 (9943)	75485 (12018)	50445 (8029)	108126 (17250)
Multiplicity	4.2 (4.2)	4.2 (4.2)	8.0 (8.1)	4.1 (3.6)
R_{sym} (%)	8.3 (50.0)	9.5 (47.9)	6.7 (42.0)	5.0 (48.6)
Completeness (%)	98.9 (96.7)	99.2 (98.2)	99.9 (99.7)	99.6 (98.7)
$\langle I/\sigma(I) \rangle$	13.9 (3.4)	9.2 (2.6)	23.1 (5.1)	16.8 (2.7)
Refinement				
Resolution range	34.9 - 1.40	29.5 - 1.32	42.3 - 1.55	36.9 - 1.17
No. of reflections used for refinement (work (free))	59973 (3157)	71710 (3775)	47922 (2523)	102719 (5407)
R_{cryst} (%)	11.6	12.1	14.5	11.3
R_{free} (%)	14.6	14.5	16.6	13.4
No. of refined residues	330	330	330	330
No. of fragment atoms	12	13	12	15
No. of other ligand atoms	34	52	43	51
No. of water molecules	329	268	294	282
RMSD, bond length (Å)	0.007	0.007	0.008	0.008
RMSD, bond angles (°)	1.2	1.2	1.2	1.3
Ramachandran plot (%) [b]				
Most favoured	93.9	93.9	93.1	93.9
Additionally allowed	6.1	6.1	6.9	6.1
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	9.0	11.4	12.4	11.1
Main chain	8.3	10.5	11.4	10.1
Side chain	9.8	12.3	13.4	12.0
Fragment atoms	15.5	24.4	20.5	10.3
Other ligand atoms [d]	22.4	27.4	35.6	26.9
Waters	23.9	23.9	25.1	24.7

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 7***

Fragment number	**109_analogue**	**112**	**114**	**125**
PDB code	**5HCO**	**4Y41**	**4Y4T**	**4Y4W**
Data collection and processing				
Wavelength (Å)	0.91841	0.8946	0.91841	0.91841
Beamline	BESSY BL14.1	BESSY BL14.3	BESSY BL14.2	BESSY BL14.2
Detector	PILATUS 6M	Mar CCD 165	MX-225 CCD	MX-225 CCD
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.2, 72.9, 52.7	45.4, 72.8, 52.8	45.3, 73.1, 52.8	45.3, 72.9, 52.6
α, β, γ (°)	90, 109.1, 90	90.0, 109.6, 90.0	90.0, 109.3, 90.0	90.0, 109.4, 90.0
Resolution range (Å)	42.8-1.17 (1.24-1.17)	42.7-1.40 (1.49-1.40)	42.7-1.30 (1.38-1.30)	42.7-1.45 (1.54-1.45)
Wilson B factor ($Å^2$)	11.4	10.7	10.4	10.5
No. of unique reflections	108391 (17012)	62976 (10055)	78761 (12429)	56566 (9000)
Multiplicity	3.7 (3.5)	4.2 (4.1)	4.1 (3.8)	4.2 (4.2)
R_{sym} (%)	3.6 (44.2)	4.3 (39.6)	5.2 (47.7)	6.3 (48.5)
Completeness (%)	99.3 (96.9)	99.6 (98.6)	99.0 (97.0)	99.1 (98.0)
$<I/\sigma(I)>$	16.8 (2.6)	22.5 (3.8)	16.4 (2.9)	16.6 (3.1)
Refinement				
Resolution range	42.8 - 1.17	29.4 - 1.40	29.5 - 1.30	28.1 - 1.45
No. of reflections used for refinement (work (free))	102971 (5420)	59827 (3149)	74823 (3938)	53737 (2829)
R_{cryst} (%)	12.5	11.5	13.3	13.6
R_{free} (%)	14.3	15.1	15.6	16.0
No. of refined residues	330	330	330	330
No. of fragment atoms	13	30	26	19
No. of other ligand atoms	31	28	41	54
No. of water molecules	289	326	314	272
RMSD, bond length (Å)	0.007	0.006	0.008	0.008
RMSD, bond angles (°)	1.2	1.2	1.3	1.2
Ramachandran plot (%) [b]				
Most favoured	93.9	94.2	94.2	93.5
Additionally allowed	6.1	5.8	5.8	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	14.1	11.9	12.4	12.0
Main chain	13.1	10.9	11.4	11.0
Side chain	15.0	12.8	13.3	13.0
Fragment atoms	15.6	19.3	16.8	22.6
Other ligand atoms [d]	38.7	26.7	30.0	31.3
Waters	31.4	28.1	25.6	24.3

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 8***

Fragment number	**131**	**158**	**162**	**164**
PDB code	**4Y4E**	**5J25**	**4Y47**	**4Y44**
Data collection and processing				
Wavelength (Å)	0.89500	0.91841	0.89440	0.91841
Beamline	BESSY BL14.3	BESSY BL14.1	BESSY BL14.3	BESSY BL14.1
Detector	Mar CCD 165	PILATUS 6M	Mar CCD 165	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.2, 52.8	45.3, 72.9, 52.6	45.3, 72.8, 52.6	45.2, 72.9, 52.6
α, β, γ (°)	90.0, 109.4, 90.0	90.0, 109.3, 90.0	90.0, 109.4, 90.0	90.0, 109.0, 90.0
Resolution range (Å)	19.2-1.30 (1.38-1.30)	42.7-1.24 (1.31-1.24)	42.7-1.19 (1.26-1.19)	49.7-1.24 (1.31-1.24)
Wilson B factor ($Å^2$)	9.1	9.9	8.6	10.1
No. of unique reflections	79467 (12595)	86046 (12396)	102282 (15746)	88968 (13388)
Multiplicity	4.1 (4.0)	3.2 (3.1)	3.8 (3.3)	3.4 (3.2)
R_{sym} (%)	6.1 (48.0)	3.0 (24.8)	6.5 (47.3)	3.2 (18.3)
Completeness (%)	99.4 (97.9)	94.2 (84.2)	98.8 (94.3)	97.5 (91.0)
$\langle I/\sigma(I) \rangle$	16.1 (3.8)	21.6 (4.1)	14.5 (3.0)	20.3 (5.3)
Refinement				
Resolution range	17.6 - 1.30	42.7 - 1.24	36.4 - 1.19	42.8 - 1.24
No. of reflections used for refinement (work (free))	75493 (3974)	81743 (4303)	97166 (5114)	84519 (4449)
R_{cryst} (%)	11.4	11.6	13.2	11.2
R_{free} (%)	13.6	13.6	15.4	12.4
No. of refined residues	330	330	330	330
No. of fragment atoms	13	26	20	10
No. of other ligand atoms	20	16	16	10
No. of water molecules	296	300	333	282
RMSD, bond length (Å)	0.007	0.007	0.006	0.006
RMSD, bond angles (°)	1.2	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	93.9	92.8	93.9	93.5
Additionally allowed	6.1	7.2	6.1	6.5
Generously allowed	0.0	0	0	0.0
Disallowed	0.0	0	0	0.0
Mean B-values ($Å^2$) [c]				
All protein atoms	10.5	11.6	10.3	11.8
Main chain	9.6	10.7	9.5	10.9
Side chain	11.4	12.4	11.0	12.6
Fragment atoms	40.2	28.5	15.5	33.1
Other ligand atoms [d]	27.4	20.5	19.9	17.9
Waters	24.5	24.7	24.0	26.5

[a] Statistic for the highest-resolution shell are shown in parenthesis; [b] Calculated by PROCHECK; [c] Calculated by MOLEMAN; [d] Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 9***

Fragment number	**171**	**181**	**189**	**203**
PDB code	**4Y3X**	**4Y3A**	**4Y3H**	**5DR0**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.9184	0.89440
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1	BESSY BL14.3
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M	Mar CCD 165
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.0, 52.7	45.2, 73.0, 52.6	45.3, 72.9, 52.6	45.2, 73.0, 52.7
α, β, γ (°)	90.0, 109.7, 90.0	90.0, 109.1, 90.0	90.0, 109.3, 90.0	90.0, 109.2, 90.0
Resolution range (Å)	42.7-1.25 (1.33-1.25)	42.7-1.17 (1.24-1.17)	42.7-1.23 (1.30-1.23)	42.7-1.40 (1.48-1.40)
Wilson B factor ($Å^2$)	8.8	9.5	9.9	9.7
No. of unique reflections	87058 (13331)	102974 (12665)	90396 (12876)	63639 (10178)
Multiplicity	3.8 (3.6)	3.5 (2.4)	3.7 (3.3)	3.9 (3.9)
R_{sym} (%)	5.9 (40.0)	3.2 (18.8)	4.0 (24.2)	6.2 (48.2)
Completeness (%)	97.5 (92.7)	94.9 (72.4)	96.3 (85.0)	99.5 (98.6)
$\langle I/\sigma(I) \rangle$	13.9 (3.0)	19.9 (4.3)	17.5 (4.3)	15.7 (3.1)
Refinement				
Resolution range	42.7 - 1.25	42.7 - 1.17	42.7 - 1.23	29.4 - 1.40
No. of reflections used for refinement (work (free))	82705 (4353)	97825 (5149)	85876 (4520)	60457 (3182)
R_{cryst} (%)	12.0	11.1	10.8	11.7
R_{free} (%)	14.2	12.9	12.8	14.3
No. of refined residues	330	330	330	330
No. of fragment atoms	26	19	9	14
No. of other ligand atoms	32	52	50	46
No. of water molecules	296	312	315	289
RMSD, bond length (Å)	0.015	0.006	0.007	0.007
RMSD, bond angles (°)	1.6	1.3	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	93.5	93.9	93.5	93.1
Additionally allowed	6.5	6.1	6.5	6.9
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	9.9	11.2	11.7	11.5
Main chain	9.1	10.2	10.8	10.6
Side chain	10.7	12.0	12.5	12.4
Fragment atoms	13.0	16.3	22.6	20.9
Other ligand atoms [d]	20.7	27.6	30.8	26.6
Waters	24.4	28.5	28.2	26.3

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 10***

Fragment number	**205**	**206**	**207**	**209**
PDB code	**4Y3L**	**4Y5A**	**4Y3T**	**5DQ5**
Data collection and processing				
Wavelength (Å)	0.89440	0.89500	0.91841	0.91841
Beamline	BESSY BL14.3	BESSY BL14.3	BESSY BL14.2	BESSY BL14.2
Detector	Mar CCD 165	Mar CCD 165	MX-225 CCD	MX-225 CCD
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.2, 72.9, 52.6	45.4, 72.9, 52.8	45.3, 72.9, 52.6	45.3, 73.0, 53.0
α, β, γ (°)	90.0, 109.5, 90.0	90.0, 109.6, 90.0	90.0, 109.4, 90.0	90, 109.5, 90
Resolution range (Å)	42.6-1.16 (1.23-1.16)	50.0-1.45 (1.48-1.45)	42.8-1.42 (1.50-1.42)	41.0-1.47 (1.56-1.47)
Wilson B factor ($Å^2$)	7.8	9.4	9.0	11.4
No. of unique reflections	108943 (15845)	57009 (2868)	60833 (9430)	54315 (8604)
Multiplicity	4.0 (3.0)	3.5 (3.4)	4.0 (3.9)	3.7 (3.4)
R_{sym} (%)	6.1 (41.9)	7.8 (47.6)	6.9 (48.5)	6 (47.5)
Completeness (%)	98.0 (88.5)	99.2 (99.9)	99.1 (95.3)	98.9 (97.5)
$\langle I/\sigma(I) \rangle$	17.7 (3.2)	16.5 (2.6)	15.7 (3.4)	16.4 (2.6)
Refinement				
Resolution range	29.4 - 1.16	26.8 - 1.45	39.5 - 1.42	36.5 - 1.47
No. of reflections used for refinement (work (free))	103495 (5447)	54088 (2897)	57791 (3042)	51599 (2716)
R_{cryst} (%)	11.6	11.9	11.7	13.0
R_{free} (%)	13.2	15.7	15.0	15.8
No. of refined residues	330	330	330	330
No. of fragment atoms	16	11	36	7
No. of other ligand atoms	60	76	72	22
No. of water molecules	298	286	289	279
RMSD, bond length (Å)	0.006	0.007	0.007	0.008
RMSD, bond angles (°)	1.2	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	93.9	93.5	94.2	93.5
Additionally allowed	6.1	6.5	5.8	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	9.3	10.1	10.2	12.3
Main chain	8.5	9.2	9.3	11.5
Side chain	10.1	11.0	11.0	13.1
Fragment atoms	17.8	15.5	20.8	30.8
Other ligand atoms [d]	22.0	32.1	30.8	24.1
Waters	24.5	25.1	26.0	24.4

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 11***

Fragment number	**211**	**216**	**218**	**224**
PDB code	**4YCK**	**4YCT**	**4YCY**	**4YD3**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841
Beamline	BESSY BL14.1	BESSY BL14.2	BESSY BL14.2	BESSY BL14.1
Detector	PILATUS 6M	MX-225 CCD	MX-225 CCD	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.0, 52.8	45.2, 73.0, 52.6	45.6, 72.9, 52.7	45.3, 73.2, 52.7
α, β, γ (°)	90.0, 109.7, 90.0	90.0, 109.5, 90.0	90.0, 108.9, 90.0	90.0, 109.7, 90.0
Resolution range (Å)	42.7-1.07 (1.13-1.07)	42.6-1.13 (1.20-1.13)	19.2-1.70 (1.80-1.70)	42.7-1.25 (1.32-1.25)
Wilson B factor ($Å^2$)	9.3	9.0	15.5	9.8
No. of unique reflections	137772 (21639)	114817 (17169)	35814 (5648)	86779 (13216)
Multiplicity	3.8 (3.7)	4.1 (3.8)	4.2 (4.2)	3.4 (3.3)
R_{sym} (%)	4.8 (54.5)	4.8 (45.1)	5.8 (49.1)	3.9 (33.1)
Completeness (%)	96.7 (94.0)	95.4 (88.5)	99.4 (98.5)	96.5 (91.0)
$\langle I/\sigma(I) \rangle$	14.2 (2.2)	17.4 (3.0)	19.5 (3.3)	18.9 (3.3)
Refinement				
Resolution range	36.5 - 1.07	34.7 - 1.13	19.3 - 1.70	42.7 - 1.25
No. of reflections used for refinement (work (free))	130883 (6889)	109076 (5741)	34023 (1791)	82440 (4339)
R_{cryst} (%)	11.9	13.8	16.6	11.1
R_{free} (%)	13.6	15.7	19.8	13.2
No. of refined residues	330	328	323	330
No. of fragment atoms	12	13	16	34
No. of other ligand atoms	32	27	19	59
No. of water molecules	302	267	164	293
RMSD, bond length (Å)	0.006	0.006	0.006	0.006
RMSD, bond angles (°)	1.2	1.2	1.0	1.2
Ramachandran plot (%) [b]				
Most favoured	93.9	93.5	93.0	94.2
Additionally allowed	6.1	6.5	7.0	5.8
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	10.7	11.1	20.4	11.4
Main chain	9.8	10.3	19.7	10.4
Side chain	11.5	11.8	21.4	12.3
Fragment atoms	12.2	16.3	38.5	15.6
Other ligand atoms [d]	19.3	27.2	35.8	23.6
Waters	25.6	24.6	27.8	27.0

[a] Statistic for the highest-resolution shell are shown in parenthesis; [b] Calculated by PROCHECK; [c] Calculated by MOLEMAN; [d] Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 12***

Fragment number	**227**	**231**	**236**	**240**
PDB code	**4YD4**	**5DR4**	**4YD5**	**4YD6**
Data collection and processing				
Wavelength (Å)	0.89440	0.89440	0.91841	0.91841
Beamline	BESSY BL14.3	BESSY BL14.3	BESSY BL14.1	BESSY BL14.1
Detector	Mar CCD 165	Mar CCD 165	PILATUS 6M	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.2, 52.7	45.2, 73.1, 52.7	45.3, 73.0, 52.7	45.3, 73.0, 52.7
α, β, γ (°)	90.0, 109.7, 90.0	90.0, 109.5, 90.0	90.0, 109.3, 90.0	90.0, 109.3, 90.0
Resolution range (Å)	42.7-1.27 (1.35-1.27)	42.6-1.50 (1.59-1.50)	42.8-1.21 (1.28-1.21)	42.6-1.30 (1.38-1.30)
Wilson B factor ($Å^2$)	10.1	11.9	9.8	11.0
No. of unique reflections	84492 (13403)	51590 (8259)	96832 (14290)	76485 (11607)
Multiplicity	4.2 (4.2)	4.2 (4.1)	3.6 (2.8)	4.2 (4.1)
R_{sym} (%)	4.8 (48.3)	7.3 (48.5)	5.2 (22.4)	6.4 (47.8)
Completeness (%)	97.8 (96.5)	99.5 (99.1)	98.1 (89.8)	96.1 (90.5)
$\langle I/\sigma(I) \rangle$	18.8 (3.2)	13.3 (3.0)	13.7 (3.6)	12.8 (2.9)
Refinement				
Resolution range	29.6 - 1.27	28.0 - 1.50	42.8 - 1.21	28.0 - 1.30
No. of reflections used for refinement (work (free))	80267 (4225)	49010 (2580)	91990 (4842)	72660 (3825)
R_{cryst} (%)	12.2	12.8	11.9	12.5
R_{free} (%)	14.7	16.0	13.6	15.1
No. of refined residues	330	330	330	330
No. of fragment atoms	10	11	13	9
No. of other ligand atoms	34	35	34	41
No. of water molecules	237	258	226	290
RMSD, bond length (Å)	0.007	0.008	0.006	0.007
RMSD, bond angles (°)	1.2	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	94.2	93.5	94.2	93.5
Additionally allowed	5.8	6.5	5.8	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	11.9	13.7	11.6	12.9
Main chain	11.0	12.9	10.6	12.0
Side chain	12.8	14.5	12.5	13.8
Fragment atoms	29.0	23.1	18.6	15.4
Other ligand atoms [d]	36.7	38.2	23.6	29.3
Waters	27.3	24.4	25.5	27.2

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 13***

Fragment number	**255**	**260**	**261**	**266**
PDB code	**4YD7**	**4Y58**	**4Y5B**	**4Y3S**
Data collection and processing				
Wavelength (Å)	0.89440	0.91841	0.91841	0.91841
Beamline	BESSY BL14.3	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
Detector	Mar CCD 165	PILATUS 6M	PILATUS 6M	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.1, 52.8	45.1, 72.9, 52.7	45.4, 73.1, 52.8	45.3, 73.1, 52.8
α, β, γ (°)	90.0, 109.6, 90.0	90.0, 109.4, 90.0	90.0, 109.5, 90.0	90.0, 109.5, 90.0
Resolution range (Å)	19.2-1.42 (1.50-1.42)	41.1-1.17 (1.24-1.17)	42.8-1.24 (1.31-1.24)	19.8-1.10 (1.17-1.10)
Wilson B factor ($Å^2$)	9.9	11.0	11.1	10.4
No. of unique reflections	61045 (9738)	106998 (16804)	91683 (14580)	127054 (19588)
Multiplicity	4.2 (4.1)	3.6 (3.4)	3.6 (3.4)	3.6 (3.4)
R_{sym} (%)	6.4 (50.2)	3.8 (48.5)	4.4 (47.3)	3.2(48.2)
Completeness (%)	99.6 (99.0)	98.7 (96.2)	99.3 (98.1)	97.0 (93.0)
$<I/\sigma(I)>$	17.3 (3.1)	17.6 (2.4)	16.7 (2.3)	20.5 (2.7)
Refinement				
Resolution range	19.2 - 1.42	39.5 - 1.17	42.8 - 1.24	19.8 - 1.10
No. of reflections used for refinement (work (free))	57994 (3051)	101648 (5350)	87098 (4585)	120701 (6353)
R_{cryst} (%)	11.6	12.6	12.1	12.4
R_{free} (%)	15.0	14.5	14.5	14.1
No. of refined residues	330	330	330	330
No. of fragment atoms	19	15	45	12
No. of other ligand atoms	39	44	24	10
No. of water molecules	307	313	264	273
RMSD, bond length (Å)	0.007	0.007	0.007	0.006
RMSD, bond angles (°)	1.2	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	93.5	93.9	93.9	93.5
Additionally allowed	6.5	6.1	6.1	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	11.3	13.1	12.9	12.7
Main chain	10.4	12.1	11.8	11.7
Side chain	12.1	13.9	13.9	13.6
Fragment atoms	18.5	21.9	35.5	42.2
Other ligand atoms [d]	33.5	28.1	26.8	17.4
Waters	25.7	27.3	27.8	27.3

[a] Statistic for the highest-resolution shell are shown in parenthesis; [b] Calculated by PROCHECK; [c] Calculated by MOLEMAN; [d] Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 14***

Fragment number	**267**	**268**	**272**	**273**
PDB code	**4Y5C**	**4Y5G**	**4Y5G**	**4Y3N**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.1, 52.8	45.4, 73.1, 52.9	45.4, 73.1, 53.9	45.6, 73.0, 53.0
α, β, γ (°)	90.0, 109.4, 90.0	90.0, 109.5, 90.0	90.0, 109.6, 90.0	90.0, 109.4, 90.0
Resolution range (Å)	42.7-1.19 (1.26-1.19)	41.2-1.12 (1.18-1.12)	42.8-1.47 (1.56-1.47)	43.0-1.34 (1.42-1.34)
Wilson B factor ($Å^2$)	11.2	10.1	11.2	12.6
No. of unique reflections	103484 (16282)	124427 (19094)	51878 (8185)	72919 (11390)
Multiplicity	3.5 (3.3)	3.6 (3.3)	2.3 (2.3)	3.6 (3.4)
R_{sym} (%)	3.7 (49.4)	4.6 (49.0)	4.9 (27.7)	4.9 (48.8)
Completeness (%)	99.1 (96.8)	98.7 (94.1)	93.5 (91.7)	99.0 (96.2)
$\langle I/\sigma(I) \rangle$	17.9 (2.2)	14.2 (2.3)	12.9 (3.5)	15.3 (2.3)
Refinement				
Resolution range	39.5 - 1.19	36.5 - 1.12	42.8 - 1.47	43.0 - 1.34
No. of reflections used for refinement (work (free))	98309 (5175)	118205 (6222)	49288 (2590)	69273 (3646)
R_{cryst} (%)	11.5	12.5	12.3	12.4
R_{free} (%)	14.1	14.2	15.6	15.2
No. of refined residues	330	330	330	330
No. of fragment atoms	16	15	18	14
No. of other ligand atoms	22	26	14	10
No. of water molecules	277	298	287	262
RMSD, bond length (Å)	0.006	0.006	0.008	0.007
RMSD, bond angles (°)	1.2	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	93.5	93.1	93.5	93.9
Additionally allowed	6.5	6.9	6.5	6.1
Generously allowed	0	0.0	0	0.0
Disallowed	0	0.0	0	0.0
Mean B-values ($Å^2$) [c]				
All protein atoms	13.2	12.3	11.8	14.5
Main chain	12.0	11.3	10.9	13.5
Side chain	14.4	13.3	12.7	15.4
Fragment atoms	34.7	24.6	17.6	24.1
Other ligand atoms [d]	26.5	25.0	19.3	24.0
Waters	27.6	26.4	26.0	28.6

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 15***

Fragment number	**274**	**Br analogue of 274**	**Cl analogue of 274**	**278**
PDB code	**4Y5K**	**5IS4**	**5ISJ**	**5DR1**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	1.00000	0.91841
Beamline	BESSY BL14.1	BESSY BL14.1	ELETTRA BL 5.2R	BESSY BL14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 2M	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.2, 52.8	45.2, 73.1, 52.7	45.1, 73.0, 52.7	45.3, 73.0, 53.0
α, β, γ (°)	90.0, 109.4, 90.0	90.0, 109.2, 90.0	90, 109.1, 90	90.0, 109.6, 90.0
Resolution range (Å)	42.7-1.44 (1.53-1.44)	42.7-1.37 (1.45-1.37)	42.7-1.65 (1.65-1.75)	42.7-1.45 (1.54-1.45)
Wilson B factor ($Å^2$)	10.9	11.9	18.0	11.1
No. of unique reflections	58762 (9404)	67872 (10813)	37526 (5699)	57303 (9151)
Multiplicity	3.7 (3.7)	3.7 (3.7)	3.4 (3.5)	3.7 (3.8)
R_{sym} (%)	5.6 (43.8)	4.4 (49.7)	3.9 (28.2)	7.1 (49.6)
Completeness (%)	99.4 (98.8)	99.4 (98.3)	96.4 (90.7)	99.2 (98.7)
$<I/\sigma(I)>$	16.9 (3.0)	17.8 (2.5)	19.4 (4.6)	13.2 (2.6)
Refinement				
Resolution range	29.5 - 1.44	42.7 - 1.37	41.1 - 1.65	36.5 - 1.45
No. of reflections used for refinement (work (free))	55826 (2936)	64478 (3394)	35649 (1877)	54437 (2866)
R_{cryst} (%)	12.1	12.9	14.0	15.5
R_{free} (%)	14.9	14.9	16.7	16.9
No. of refined residues	330	330	330	330
No. of fragment atoms	11	11	11	18
No. of other ligand atoms	30	31	31	38
No. of water molecules	283	283	253	236
RMSD, bond length (Å)	0.007	0.01	0.009	0.008
RMSD, bond angles (°)	1.2	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	93.1	94.2	93.1	93.5
Additionally allowed	6.9	5.8	6.9	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	12.1	16.3	21.8	12.8
Main chain	11.1	15.2	20.2	11.7
Side chain	13.0	17.2	23.4	13.8
Fragment atoms	15.0	21.1	29.6	31.5
Other ligand atoms [d]	27.2	38.0	46.5	28.1
Waters	27.0	29.2	32.5	22.0

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 16***

Fragment number	**285**	**286**	**290**	**291**
PDB code	**4Y3G**	**4Y4A**	**4Y35**	**4Y45**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.3, 73.1, 52.8	45.3, 73.1, 52.8	45.3, 73.2, 52.8	45.3, 73.0, 52.8
α, β, γ (°)	90.0, 109.5, 90.0	90.0, 109.6, 90.0	90.0, 109.5, 90.0	90.0, 109.6, 90.0
Resolution range (Å)	19.8-1.13 (1.20-1.13)	41.2-1.28 (1.36-1.28)	41.1-1.46 (1.55-1.46)	42.7-1.06 (1.12-1.06)
Wilson B factor ($Å^2$)	9.8	10.7	12.2	9.7
No. of unique reflections	117892 (18348)	83422 (13296)	56120 (8860)	144991 (22289)
Multiplicity	3.7 (3.6)	3.7 (3.6)	3.8 (3.8)	3.6 (3.3)
R_{sym} (%)	6.4 (47.5)	4.5 (42.7)	6.2 (44.2)	3.8 (46.5)
Completeness (%)	97.5 (94.3)	99.5 (98.2)	99.3 (97.8)	98.8 (94.1)
$\langle I/\sigma(I) \rangle$	11.2 (2.5)	16.5 (2.9)	13.3 (3.4)	16.7 (2.4)
Refinement				
Resolution range	19.8 - 1.13	36.6 - 1.28	39.6 - 1.46	42.7 - 1.06
No. of reflections used for refinement (work (free))	111997 (5895)	79251 (4171)	53313 (2807)	137741 (7250)
R_{cryst} (%)	16.3	11.6	12.5	12.0
R_{free} (%)	18.4	13.9	15.9	13.6
No. of refined residues	330	330	330	330
No. of fragment atoms	12	11	12	31
No. of other ligand atoms	6	16	24	18
No. of water molecules	273	306	270	288
RMSD, bond length (Å)	0.007	0.006	0.007	0.006
RMSD, bond angles (°)	1.2	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	93.5	93.9	93.5	92.8
Additionally allowed	6.5	6.1	6.5	7.2
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	11.3	12.7	13.8	11.5
Main chain	10.5	11.6	13.0	10.5
Side chain	12.1	13.8	14.6	12.4
Fragment atoms	20.1	17.4	21.6	16.1
Other ligand atoms [d]	16.4	23.2	30.3	16.2
Waters	24.8	27.3	27.3	24.9

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 17***

Fragment number	**305**	**306**	**311**	**321**
PDB code	**4Y37**	**4Y3R**	**5DR7**	**4Y3D**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.89500	0.89440
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.3	BESSY BL14.3
Detector	PILATUS 6M	PILATUS 6M	Mar CCD 165	Mar CCD 165
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.4, 73.1, 52.9	45.2, 73.2, 52.7	45.3, 73.1, 52.8	45.2, 72.9, 52.6
α, β, γ (°)	90.0, 109.4, 90.0	90.0, 109.6, 90.0	90, 109.4, 90	90.0, 109.4, 90.0
Resolution range (Å)	19.2-1.69 (1.79-1.69)	42.6-1.13 (1.20-1.13)	42.8-1.23 (1.31-1.23)	42.7-1.48 (1.57-1.48)
Wilson B factor ($Å^2$)	13.7	10.3	8.2	10.8
No. of unique reflections	36243 (5481)	120427 (19326)	89949 (13347)	53444 (8611)
Multiplicity	3.7 (3.3)	3.7 (3.5)	4.2 (4.1)	4.2 (4.1)
R_{sym} (%)	9.9 (48.5)	4.1 (49.9)	9.6 (46.1)	6.9 (53.9)
Completeness (%)	98.4 (93.4)	99.6 (99.2)	96.5 (89.1)	99.7 (99.6)
$<I/\sigma(I)>$	12.1 (2.5)	16.7 (2.3)	11.9 (3.3)	16.2 (2.7)
Refinement				
Resolution range	19.2 - 1.69	28.0 - 1.13	25.5 - 1.23	28.1 - 1.48
No. of reflections used for refinement (work (free))	34429 (1812)	114405 (6022)	85451 (4498)	50771 (2673)
R_{cryst} (%)	18.1	12.4	12.3	11.7
R_{free} (%)	21.7	14.1	14.5	15.7
No. of refined residues	330	330	330	330
No. of fragment atoms	11	12	17	16
No. of other ligand atoms	10	20	44	24
No. of water molecules	268	290	289	320
RMSD, bond length (Å)	0.007	0.006	0.009	0.007
RMSD, bond angles (°)	1.0	1.2	1.3	1.2
Ramachandran plot (%) [b]				
Most favoured	93.1	93.1	93.9	93.5
Additionally allowed	6.9	6.9	6.1	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	15.6	12.4	10.0	12.1
Main chain	15.0	11.4	9.1	11.2
Side chain	16.4	13.3	10.9	12.9
Fragment atoms	23.9	15.8	24.3	17.4
Other ligand atoms [d]	27.8	21.9	29.5	27.8
Waters	26.9	27.7	22.6	28.3

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 18***

Fragment number	**323**	**328**	**330**	**333**
PDB code	**4Y4B**	**4Y5M**	**5DR8**	**5DR3**
Data collection and processing				
Wavelength (Å)	0.89500	0.89500	0.895	0.89500
Beamline	BESSY BL14.3	BESSY BL14.3	BESSY BL14.3	BESSY BL14.3
Detector	Mar CCD 165	Mar CCD 165	Mar CCD 165	Mar CCD 165
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.4, 73.2, 52.9	45.3, 72.8, 52.9	45.3, 73.2, 52.9	45.3, 73.0, 52.8
α, β, γ (°)	90.0, 109.5, 90.0	90.0, 109.9, 90.0	90.0, 109.6, 90.0	90.0, 109.5, 90.0
Resolution range (Å)	42.8-1.11 (1.18-1.11)	42.6-1.28 (1.36-1.28)	42.7-1.47 (1.56-1.47)	42.7-1.24 (1.31-1.24)
Wilson B factor ($Å^2$)	10.1	10.4	8.8	9.6
No. of unique reflections	125728 (18023)	82855 (13281)	55239 (8809)	91107 (14458)
Multiplicity	3.9 (3.1)	4.1 (4.1)	4.2 (4.1)	4.2 (4.1)
R_{sym} (%)	4.8 (49.6)	5.0 (57.0)	7.1 (48.7)	4.4 (43.3)
Completeness (%)	97.4 (86.7)	99.5 (99.1)	99.6 (98.8)	99.2 (97.8)
$\langle I/\sigma(I) \rangle$	14.3 (2.3)	18.5 (2.5)	19.3 (4.4)	20.3 (3.3)
Refinement				
Resolution range	28.2 - 1.11	25.6 - 1.28	28.1 - 1.47	20.9 - 1.24
No. of reflections used for refinement (work (free))	119441 (6287)	78712 (4143)	52476 (2763)	86551 (4556)
R_{cryst} (%)	12.5	11.9	12.3	11.2
R_{free} (%)	14.2	14.6	15.1	13.3
No. of refined residues	330	330	330	330
No. of fragment atoms	11	31	11	14
No. of other ligand atoms	36	16	20	28
No. of water molecules	290	366	274	341
RMSD, bond length (Å)	0.006	0.007	0.008	0.006
RMSD, bond angles (°)	1.2	1.2	1.2	1.2
Ramachandran plot (%) [b]				
Most favoured	93.5	93.1	93.5	93.5
Additionally allowed	6.5	6.9	6.5	6.5
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	12.7	12.4	9.7	12.0
Main chain	11.8	11.5	9.0	11.0
Side chain	13.5	13.3	10.5	12.9
Fragment atoms	45.8	27.7	30.8	17.5
Other ligand atoms [d]	32.8	23.4	32.6	29.1
Waters	29.3	29.4	21.6	26.6

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

Table 8.1 *Data collection and refinement statistics of fragments –* ***part 19***

Fragment number	**337**	**338**
PDB code	**4Y5N**	**4Y5P**
Data collection and processing		
Wavelength (Å)	0.89440	0.89440
Beamline	BESSY BL14.3	BESSY BL14.3
Detector	Mar CCD 165	Mar CCD 165
Space group	$P2_1$	$P2_1$
Unit-cell parameters [a]		
a, b, c (Å)	45.3, 72.9, 52.8	45.3, 72.9, 52.7
α, β, γ (°)	90.0, 109.6, 90.0	90.0, 109.7, 90.0
Resolution range (Å)	42.6-1.29 (1.37-1.29)	42.7-1.23 (1.30-1.23)
Wilson B factor ($Å^2$)	8.7	9.5
No. of unique reflections	79030 (12449)	92370 (13924)
Multiplicity	4.3 (4.3)	4.1 (3.3)
R_{sym} (%)	6.3 (49.9)	5.2 (48.1)
Completeness (%)	97.2 (95.1)	98.5 (92.2)
$\langle I/\sigma(I) \rangle$	16.6 (3.5)	16.6 (2.6)
Refinement		
Resolution range	42.6 - 1.29	25.5 - 1.23
No. of reflections used for refinement (work (free))	75077 (3953)	87751 (4619)
R_{cryst} (%)	12.1	12.2
R_{free} (%)	14.5	14.7
No. of refined residues	330	330
No. of fragment atoms	19	16
No. of other ligand atoms	20	26
No. of water molecules	307	312
RMSD, bond length (Å)	0.007	0.007
RMSD, bond angles (°)	1.2	1.2
Ramachandran plot (%) [b]		
Most favoured	92.8	93.5
Additionally allowed	7.2	6.5
Generously allowed	0	0
Disallowed	0	0
Mean B-values ($Å^2$) [c]		
All protein atoms	10.0	11.3
Main chain	9.3	10.5
Side chain	10.8	12.2
Fragment atoms	15.3	19.1
Other ligand atoms [d]	18.2	22.4
Waters	24.6	25.3

[a] *Statistic for the highest-resolution shell are shown in parenthesis;* [b] *Calculated by PROCHECK;* [c] *Calculated by MOLEMAN;* [d] *Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.*

8.2 Data collection and refinement statistics of DCC-products

Table 8.2 *Data collection and refinement statistics: DCC project.*

Fragment number	**20 mM acylhydrazone**	**100 mM acylhydrazone**	**2 mM acylhydrazone**	**bis-1**
PDB code	**4KUP**	**4LBT**	**4LHH**	**5HCT**
Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841
Beamline	BESSY BL14.2	BESSY BL14.2	BESSY BL14.2	BESSY BL14.1
Detector	MX-225 CCD	MX-225 CCD	MX-225 CCD	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters [a]				
a, b, c (Å)	45.4, 72.7, 52.8	45.4, 72.9, 52.9	45.5, 73.1, 53.1	45.2, 72.8, 52.7
α, β, γ (°)	90.0, 109.6, 90.0	90.0, 109.6, 90.0	90.0, 109.3, 90.0	90.0, 108.9, 90.0
Resolution range (Å)	30.0-1.31 (1.33-1.31)	30.0-1.25 (1.27-1.25)	30.0-1.73 (1.76-1.73)	42.8-1.36 (1.44-1.36)
Wilson B factor ($Å^2$)	10.8	9.8	13.6	12.0
No. of unique reflections	76831 (3792)	85603 (3157)	34335 (1708)	69237 (11019)
Multiplicity	4.1 (3.9)	4.2 (3.4)	4.1 (4.0)	3.8 (3.6)
R_{sym} (%)	5.1 (40.1)	5.1 (34.5)	8.4 (48.5)	6.6 (49.1)
Completeness (%)	99.0 (97.3)	95.4 (70.4)	99.8 (99.1)	99.6 (98.8)
$\langle I/\sigma(I) \rangle$	25.7 (3.7)	26.5 (3.3)	17.4 (2.5)	11.3 (2.4)
Refinement				
Resolution range	26.8 - 1.31	29.5 - 1.25	26.3 - 1.73	42.8 - 1.36
No. of reflections used for refinement (work (free))	72943 (3860)	81291 (4309)	32583 (1734)	65775 (3462)
R_{cryst} (%)	13.0	12.2	14.6	12.1
R_{free} (%)	14.9	13.9	17.5	15.0
No. of refined residues	330	330	330	330
No. of fragment atoms	38	68	26	24
No. of other ligand atoms	48	16	44	49
No. of water molecules	335	367	333	234
RMSD, bond length (Å)	0.008	0.005	0.006	0.007
RMSD, bond angles (°)	1.2	1.1	1.0	1.2
Ramachandran plot (%) [b]				
Most favoured	94.2	93.9	94.2	93.1
Additionally allowed	5.8	6.1	5.8	6.9
Generously allowed	0	0	0	0
Disallowed	0	0	0	0
Mean B-values ($Å^2$) [c]				
All protein atoms	11.6	11.0	12.7	14.4
Main chain	11.1	10.4	12.0	13.4
Side chain	12.3	11.6	13.6	15.3
Fragment atoms	18.6	15.1	16.5	18.5
Other ligand atoms [d]	20.8	13.3	29.4	31.5
Waters	26.1	26.5	27.5	29.1

[a] Statistic for the highest-resolution shell are shown in parenthesis; [b] Calculated by PROCHECK; [c] Calculated by MOLEMAN; [d] Other ligands are glycerol, DMSO, EG, PEG, 1PE, and acetate. Columns are sorted according to the order the fragments are mentioned in the text.

Abbreviations List

Å	Angström (1 Å = 10^{-10} m)
ACE	angiotensin converting enzyme
ACT	artemisinin-based combination therapy
ADME parameters	Adsorption, distribution, metabolism, excretion
AIDS	Acquired immunodeficiency syndrome
Ala	Alanine
APP	ß-amyloid precursor protein
Asp	Aspartate
Aß	Amyloid-ß protein
BACE1	ß-site amyloid precursor protein cleaving enzyme 1
BB	Building block
CCD camera	Charge-coupled device camera
clogP	Calculated logarithm of the octanol-water partition coefficient
CTL [%]	Percentage control values
Cys	Cysteine
Da	Dalton (1Da = 1g/mol)
DCC	Dynamic combinatorial chemistry
DCL	Dynamic combinatorial library
DMSO	Dimethylsulfoxide
EP	Endothiapepsin
ESI-MS	Electrospray Ionization mass spectrometry
FBDD	Fragment-based drug discovery
FBLD	Fragment-based lead discovery
F_c	Calculated structure factor amplitudes
FDA	US Food and Drug Administration
F_{hkl}	Structure factor
$\|F_{hkl}\|$	Structure factor amplitude
$F_{hkl} = F_{-h-k-l}$	Friedel's law
F_{hot}	Fluorescence decrease
$F_{initial}$	Initial fluorescence
F_{norm}	Normalized fluorescence
F_o	Observed structure factor amplitudes
Gly	Glycine
GOL	glycerol
ΔG^0	Gibbs free energy
HA	Heavy atoms
h, *k*, *l*	Miller indices

H-bond	Hydrogen bond
HIV1 protease	human immunodeficiency virus type 1 protease
HPLC–MS = LC-MS	High-performance liquid chromatography-mass spectrometry
HTS	High-throughput screening
I*(hkl)*	Intensity of the reflection spots
IC_{50}	Half maximal inhibitory concentration
Ile	Isoleucine
IR-Laser	Infrared Laser
ITC	isothermal titration calorimetry
K	Kelvin
K_d	Dissociation constant
LE	Ligand efficiency
Leu	Leucine
µ	micro-
m	milli-
M	Molarity ($mol \cdot l^{-1}$)
MAD	Multiple-wavelength anomalous dispersion
MHz	Megahertz
MIR	Multiple isomorphous replacement
MR	Molecular replacement
MST	Microscale Thermophoresis
MW	Molecular weight
NaAc	Sodium acetate
NaCl	Sodium chloride
NaN3	Sodium azide
NH_4Ac	Ammonium acetate
NMR	Nuclear magnetic resonance
NSAID	Nonsteroidal anti-inflammatory drug
o/w	Oil in water
occ.	Occupancy
PBS	Phosphate-buffered saline
pD	Negative decimal logarithm of the deuterium ion activity
PDB	Protein data bank
PEG	Polyethylene glycol
pH	Negative decimal logarithm of the hydrogen ion activity
Phe	Phenylalanine
pI	Isoelectric point
pK_a	Negative decimal logarithm of the acid dissociation constant
RDA	Reporter displacement assay
RO3	Rule of three
RO5	Lipinski's rule of five

R_{sym}	Agreement between symmetry-related reflections
SAD	Single-wavelength anomalous dispersion
SAP	Secreted aspartic proteases
SAR by NMR	Structure-activity relationship by nuclear magnetic resonance
SBDD	Structure-based drug design
SeMet	Selenomethionine
Ser	Serine
SPR	Surface Plasmon resonance
S_T	Soret coefficient
STD-NMR	Saturation transfer difference nuclear magnetic resonance
Thr	Threonine
T_m	Melting temperature
TPSA	Total polar surface area
TSA	Thermal shift assay
Tyr	Tyrosine
u, *v*, *w*	Patterson cell coordinates
UNAIDS	United Nations Programme on HIV/AIDS
v/v	Volume per volume
w/v	Weight per volume
WHO	World health organization
x, *y*, *z*	Real cell coordinates
φ(hkl)	Phase angle

Acknowledgments / Danksagung

Prof. Dr. Gerhard Klebe danke ich herzlich für die unermüdliche Betreuung und das Vertrauen, das er in mich setzte als er mich in seine Gruppe aufnahm. Ich bedanke mich für die interessante Themenstellung dieser Doktorarbeit sowie für seine stets konstruktiven Vorschläge bei der Durchführung und Auswertung verschiedenster Aufgaben. Ich danke ihm dafür, dass er mich stets gefördert hat und mir dadurch die Präsentation meiner Arbeit auf nationalen und internationalen Tagungen ermöglicht hat. Darüber hinaus danke ich ihm auch dafür, dass er trotz maximaler Auslastung immer auch Zeit für persönliche Gespräche fand, die einem in Momenten der Unsicherheit weitergeholfen haben. Es war mir in jeder Hinsicht eine Ehre und größtes Vergnügen von ihm zu lernen und für ihn zu arbeiten.

Prof. Dr. Andreas Heine gebührt mein Dank für die ausführliche Einführung in das Gebiet der Röntgenkristallographie. Ich danke ihm für seine Geduld und die vielen Stunden, die er mit mir an zahlreichen Datensätzen verbracht hat. Ganz besonders danke ich ihm für seinen riesigen Einsatz in dem Endothiapepsinprojekt, das ohne seine Hilfe und Unterstützung unmöglich erfolgreich beenden worden wäre. Darüber hinaus danke ich ihm für das Korrekturlesen der Manuskripte und für seine Bereitschaft meine Doktorarbeit als Zweitgutachter zu bewerten.

Dem Endothiapepsinteam, und ganz besonders *Dr. Johannes Schiebel, Frederik Ehrmann, Stefan Krimmer, Martin Stieler, Kann Fu* und *Xiaojie Wang* danke ich sehr für die enorme Unterstützung bei der Datensammlung und Verfeinerung der über 450 Datensätze. Darüber hinaus danke ich *Dr. Johan Winquist*, der mir bei der Vorbereitung der Proteinkristalle und der Soakingexperimente besonders am Anfang des Projektes sehr geholfen hat. Ohne euren Einsatz wäre es mir kaum möglich gewesen all die Arbeit zu bewältigen. Danke!

Dr. Nina Schlink und *Dr. Astrid Sitte* (NanoTemper Technologies) sowie *Dr. Markus Zeeb* und *Franziska Popp* (Boehringer Ingelheim) danke ich sehr für ihre Mitwirkung bei der Etablierung und Durchführung des MST-Assays.

Prof. Dr. Anna Hirsch und *Dr. Milon Mondal* von der *Universität in Groningen* danke ich für die erfolgreiche Zusammenarbeit im Bereich der dynamisch-kombinatorischen Chemie.

Dr. Uwe Müller, *Dr. Manfred Weiss* und *Dr. Franziska Huschmann* vom *Helmholz-Zentrum für Materialien und Energie in Berlin* danke ich für die hilfreiche Unterstützung bei der Datensammlung am Synchrotron.

Ein riesen Dankeschön an Frau *Lydia Hartleben* für ihre hervorragende Unterstützung bei der Organisation aller bürokratischen Abläufe.

Herrn *Christian Sohn* danke ich für die technische Unterstützung und seine Hilfe an dem MAR-Röntgengerät.

Dr. Johannes Schiebel danke ich für die erfolgreiche Kooperation und sein Engagement in dem Endothiapepsinprojekt. Sein außerordentlich breites Wissen und sein Fleiß haben mich immer wieder inspiriert. Danke dafür!

Dr. Alexander Metz bin ich für das Korrekturlesen der Manuskripte und für seine Art „über den Tellerrand hinauszuschauen“ sehr dankbar. Ich danke ihm für die vielen konstruktiven Vorschläge und dafür, dass er stets für gute Laune in der Arbeitsgruppe gesorgt hat.

Dr. Johan Winquist danke ich sehr für die Zusammenarbeit am Anfang meiner Promotionszeit. Ich danke ihm dafür, dass er mir nicht nur viele Grundlagen beigebracht hat sondern mir auch eine enorme Hilfe bei der Optimierung meiner Englischkenntnisse war. Seine frohe Natur, die Lust auf Musik und die vielen Gespräche haben stets für eine sehr angenehme Arbeitsatmosphäre gesorgt. Tack så mycket!

Für die erfolgreiche Zusammenarbeit und die immer entspannte Arbeitsatmosphäre bei der Betreuung und Ausbildung der Pharmaziestudenten im 4. Semester danke ich meinen Kollegen von der *AG Keusgen*, ganz besonders Herrn *Floris van Elsäcker* und *Dr. Doru Vornicescu*.

Meinen *Bürokollegen* möchte ich für die angenehme Arbeitsatmosphäre und die vielen Gespräche danken, im Besonderen *Jonathan, Angela, Lorena, Jessi, Christina, Janis, Steffi* und *Tanja.* Ich werde euch vermissen!

Darüber hinaus danke ich allen fleißigen Mensagängern, vor allem *Martin, Manuel, Frederik, Phong* und *Chris* für die gemeinsam verbrachten Mittagsstunden.

Meinen Freunden aus Studienzeiten, *Eggert, Chris, Katharina, Frederik*, *Katja, Bastian*, *Kim, Eva*, *Christina, Phong, Sascha, David* und *Heiko* danke ich für die häufigen Treffen, die Skiausflüge und die tolle Zeit in und außerhalb von Marburg. Besonders danke ich *Katharina & Frederik,* die mir seit über 10 Jahren treue Freunde sind und mit denen wir Vieles gemeinsam erlebt haben. Möge diese Freundschaft nie enden!

Ein ganz besonderer Dank geht an meine liebe *Frau Stuhl*, die mir in den letzten sechs Jahren beste Freundin, Schwester und Mutter zugleich war. Ich werde ihr für immer dafür dankbar sein, dass sie mich bei sich aufgenommen hatte, in der Lernphase für das zweite Staatsexamen oft für mich gekocht und für die Prüfungen Woche für Woche mitgefiebert hatte. Für all die guten Zeiten und die vielen Gespräche bin ich sehr dankbar und überglücklich, dass ich sie kennenlernen durfte. Sie wird für immer in meinem Herzen bleiben!

Mein tiefster Dank gebührt meiner großartigen Familie, insbesondere meinen Eltern, *Tanya* und *Georg*, meiner Schwester und meinem Schwager, *Gergana* und *Yanko*, sowie meinem Opa *Mitko* für ihre grenzlose Unterstützung und Liebe in all den Jahren. Ohne euch wäre ich nie so weit gekommen! Ihr macht mich unglaublich stolz und ich bin überglücklich, dass ich euch habe!

Der letzte und größte Dank ergeht an meinen Partner, *Dr. Eggert Rühmann*, der mich mit meinem südländischen Temperament seit über acht Jahren aushält☺. Ich danke ihm dafür, dass er mir während des Studiums und der Promotion ein unermüdlicher Diskussionspartner war, viele Korrekturen erledigt und mich stets motivierend begleitet hat. Danke für all seine Unterstützung, Vertrauen und Liebe. ♥

ERKLÄRUNG

Ich versichere, dass ich meine Dissertation

“From Active-Site Mapping to Lead Discovery using Fragment-based Approaches on the Aspartic protease Endothiapepsin”

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen sind Zitate als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den

.....

(Nedyalka Radeva)

Curriculum vitae

Aus Gründen des Persönlichkeitsschutzes wird von der elektronischen Veröffentlichung des Lebenslaufes abgesehen.